Spring 5-15-2020

Understanding Task Interference in Prospective Memory Using On-Line Probes: Strategic Delay or Limited-Capacity Monitoring?

Francis T. Anderson
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychological and Brain Sciences

Dissertation Examination Committee:
Mark A. McDaniel (Chair)
David A. Balota
Julie M. Bugg
Erin R. Foster
Michael J. Strube

Understanding Task Interference in Prospective Memory Using On-Line Probes: Strategic Delay or Limited-Capacity Monitoring?

by

Francis T. Anderson

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2020
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

Grad school is a rough time for everyone; there is no doubt about that. Everyone has their reasons why – some people don't like the work, some want to start a family, others may have financial concerns – but guaranteed, no one has ever said, "Oh, earning my PhD? Yeah, I mean it took awhile, but it wasn't so difficult."

But when I actually pause to think about the sheer number of people, and the amount of time they put in to get me to this point, my slight difficulties seem kinda paltry. The beginning, middle, and end of my acknowledgments are of course reserved for my mother, saint that she is. I can't say enough, and so I won't really try: If you're looking for someone to blame for my existence, look no further than Katherine Anderson, AKA mama. The support that she – and the rest of my family: her husband Craig, my grandmother Frances, my brothers Calhoun and William – have given me implicitly and explicitly are the only reason I haven't given up numerous times in grad school and in life.

I also can't say enough about the primary education I've received. I attended some pretty phenomenal public schools, in a state where that is not the norm, and so I have to show some love to Brennan Elementary, Crayton Middle, and Dreher High. Special shout-outs to Mrs. Lumpkin, Mrs. Williams and Mrs. Murphy, Mr. Vandiver, Mrs. Stover, Mrs. Sunday, and Mrs. Peterson, who are such wonderful educators and people.

At my undergrad, Furman, it's no secret that I didn't get along or fit in. I was miserable, and ready to graduate a year early with my degree in philosophy so I could just escape. But, one of the major reasons I stuck around was Gil Einstein, the professor who introduced me to science, psychology, and cognition more specifically. One day, toward the end of my spring semester Intro Psych course, Gil approached me and said, "I like your brain." Of course, he

subsequently clarified, “I like the way you think. How would you like to get involved in some research?” He then kept tabs on me to make sure I didn’t completely bomb Research Methods and Stats, and after that I was part of the lab. There is no understating the influence Gil had on my future; I can with 99% confidence say that I would not be doing anything related to psychology, and maybe not even science, if it wasn’t for him. Quite frankly, I could easily be tending bar somewhere right now with my less-than-useful philosophy degree. The rest of Furman’s psychology department, particularly John Batson and Michelle Horhota, were also incredibly supportive and encouraging. I can’t say enough good things about that whole department. They’re educating the *right* way, with a great culture fostered by Gil and John and [the legendary] Charles Brewer. Finally, shout-out to my only real friends at Furman (outside of the gym, that is), Lynn and Mark Roosevelt. They are incredibly giving, always have a refreshing perspective, and we have remained close through grad school.

Remove one of these people from my life and I’m pretty sure the whole thing would be turned on its head, and at this point I haven’t even made it to WashU yet. First and foremost, my advisor and mentor Mark McDaniel has had my back and encouraged my growth throughout. He has taught me so much about how to think, write, and do research that it is ridiculous; I couldn’t have asked for a mentor more closely suited to my personality and work style, or one that genuinely cares about me independent from my research productivity. I know some of this could be hindsight, but I seriously believe that pairing me with any other professor would have been a step down in terms of compatibility, and I think (hope) Mark agrees. It’s been a good ride, Mark. My other committee members, too, each have played their own role in my development. I had the good fortune of working with Erin on a prospective memory grant, I’ve picked Dave’s brain about diffusion models on a couple of occasions, I’ve taken every stats class Mike has offered, and Julie has critiqued and supported me as a committee member for my Masters, Subject Matter, and now Dissertation.

Equally important are my friends, who have walked through it with me hand in hand. My lab mates Toshi Miyatsu and Reshma Gouravajhala have been there whenever I've needed it with a quip, solid advice, or a kick in the rear. Usually all three, and in that order. Well, Reshma sometimes just skips step 2. They have absolutely been the glue that has held me together during grad school, and will always remain dear friends. If Reshma and Toshi were the glue, then my accomplice Öykü Üner was the person who picked up the pieces in the first place. She's an exceptional person who somehow always takes the bad with the good, sees the best side of others, and has an overflowing well of patience. I like to think that, when combined, my impatience speeds up her decision-making process—and there is no doubt I'm now making fewer errors than I was before. By the way, for those who didn't get it, that was a (really bad) accumulator model joke. Everyone, collectively now, *sigh*.

Francis Anderson

Washington University in St. Louis

May 2020

Dedicated to Katherine, Frances, and Rose, the three women who – for better or worse :) – raised me into who I am today.

Abstract of the Dissertation

Understanding Task Interference in Prospective Memory Using On-Line Probes:

Strategic Delay or Limited-Capacity Monitoring?

by

Francis Anderson

Doctor of Philosophy in Psychological & Brain Sciences

Washington University in St. Louis, 2020

Professor Mark A. McDaniel, Chairperson

In prospective memory (PM) research, a common finding is that people are generally slower to respond to a given ongoing task (OT) when they have to perform a PM task concurrently, as compared to performing the OT alone. Multiprocess theory claims that this slowing, termed task interference, is indicative of monitoring processes. Monitoring is thought to be cognitively demanding and heavily reliant on working memory, as people hold their intention in mind and look for features relevant to the PM task. PM decision control (PMDC) theory, instead, proposes that task interference reflects a strategic and intentional delay strategy. To address this theoretical dispute, in the present study we first develop and validate a within-block probe procedure that measures self-reported difficulty and motivation, in order to determine their relationship with task interference. Monitoring could involve motivational factors—in that some people may not care enough to monitor—but clearly implicates difficulty as a fundamental correlate of task interference. By contrast, the PMDC model clearly excludes difficulty as having any association with task interference, but one's motivation to perform well on the PM task could potentially explain differences in task interference. After validating the probe procedure in the first

experiment, we applied the methodology in two following experiments designed to manipulate task interference, with the central question being, “Is task interference (i.e., change in behavior due to PM) associated with an increase in self-reported difficulty, or is it better tracked by self-reported motivation to perform the PM task?” In both experiments, we found that difficulty ratings – not motivation – was consistently related to task interference, supporting a multiprocess view.

Chapter 1: Introduction

Prospective memory (PM) is an individual's ability to execute an intended action at a later time or event. Characteristically, a PM task is embedded within ongoing activity, requiring someone to stop whatever they happen to be doing at the appropriate moment and execute their intention. For example, you might form the intention to make a phone call on the walk to work in the morning, requiring you to appropriately remember the intention at the beginning of the walk, rather than getting distracted by the usual morning bustle. This example highlights another unique feature of PM—the large degree of self-initiation. Successful PM retrieval not only requires retrospectively remembering the content of the intention, but also prospectively self-initiating the intended action without any external prompt. Retrospective memory, by contrast, is often externally requested (e.g., a friend asking how your day has been), but that is far more rare with PM. Even people that religiously update their smart-phone calendars, thereby offloading the prospective component onto external reminders, must remember to actually put the intention in their phone to begin with, and must also remember to perform numerous other intentions most people do not set reminders for (e.g., vacuum the floor, get groceries, tweak that statistical analysis).

In the laboratory, PM is often studied by giving participants an intention, such as pressing the keyboard in response to a particular cue, also called a target, which is embedded within the context of an ongoing task (OT). In this scenario, for some of the OT trials the stimulus is actually a PM target. For example, given an ongoing 2-back task, where participants decide whether the word seen two trials previously was the same as the word currently on the screen, and the PM intention to press the *Q* key whenever a string of letters appears containing the consecutive letters *tor* (e.g., sena*tor*), participants must recognize while performing the 2-back task that the stimulus happens to also be a target (i.e., it contains *tor*). Successful detection is measured by appropriately suspending the OT response (i.e., yes or no) in favor of making a PM response (i.e., press the *Q* key). Some researchers will also let participants make the PM response on the next couple of trials, which stresses retrieval of the intention over the exact timing of the action in its measure of PM accuracy. Many other variables are also in the hands of the researchers, such as the total number of PM and OT trials, where the PM trials occur during the experiment, and of course the exact PM task instructions.

This paradigm (Einstein & McDaniel, 1990) has been used extensively in the PM literature, not only because of its simplicity, but also its real-world applicability (Dismukes, 2012). Just as in the real world, ongoing activity is the primary task for the participant, addressing demands and making judgments as needed. The PM task is embedded within the OT, therefore requiring someone to recognize the significance of the PM cue, interrupt whatever they happened to be doing at the time, and perform the intended action. For instance, given the intention to get gas on the way home from work, one has to stop dodging dangerous drivers (the OT) and remember to take the appropriate turn (the cue) toward the gas station, instead of going straight home as usual.

A particularly influential application of the paradigm has attempted to isolate the processes involved in PM retrieval by examining OT behavior both with and without possessing an additional PM intention. During the *control block*, participants perform the OT alone, serving as a standard of comparison for the *PM block*, where participants perform the exact same OT but now also have a PM task. When comparing these two blocks (or between-subjects groups) a common finding is that participants are slower to respond to the OT in the PM block than in the control block. This OT response time slowing, termed *task interference* or cost, therefore must in some way be related to performing the PM intention, operating under the assumption that the response time cost reflects additional PM-related processing (see Smith, 2003).

Chapter 2: Theoretical Accounts of Task Interference

2.1 Monitoring

To account for PM-related task interference, Smith (2003) proposed the PAM (preparatory attentional and memory) theory of PM retrieval, which contends that task interference signifies the engagement of preparatory attentional processes to support cue detection. Smith describes these preparatory processes (hereafter described as *monitoring*) as being characterized by "some level of nonautomatic monitoring of the environment for the occurrence of prospective memory target events" (Smith, 2003, p. 349) and, "may include rehearsal of the prospective memory target events" (Smith & Bayen, 2004, p. 757). A critical assumption of PAM theory, however, is that attentional resources must be engaged throughout the OT to ensure detection; if any PM target is presented during a lapse in monitoring, then it follows that the stimulus will only be evaluated according to the OT, and the target will be missed. Around the same time, Guynn (2003) proposed a two-process theory of task interference, which states that people must constantly remain in a "retrieval mode," maintaining high activation of the PM task, and on each trial they search for features relevant to performing the PM task. Though similar in many respects to PAM theory, Guynn's model explicitly attributes task interference to two distinct behaviors: a sustained retrieval mode followed by target-checking. As is clear from both theories, a key characteristic of monitoring is that it is a limited-capacity cognitive process, meaning that monitoring draws on attentional resources, is demanding and challenging, and reliant on working memory.

The multiprocess theory (McDaniel & Einstein, 2000; 2007) differs from PAM theory by proposing an additional route to PM retrieval, other than monitoring. Research has shown that participants can sometimes achieve high PM accuracy in the absence of task interference—which is problematic for PAM theory because it implies participants were able to successfully retrieve the intention without monitoring (Einstein, McDaniel, Thomas, Mayfield, Shank, Morrisette, & Breneiser, 2005; Harrison & Einstein, 2010; Scullin, McDaniel, Shelton, & Lee, 2010b; Scullin, McDaniel, & Einstein, 2010a). Based on these studies and others, McDaniel and Einstein theorized that successful PM can also occur when a strong external cue triggers *spontaneous retrieval* of the intention. Often phenomenologically described as the intention, "popping into mind," spontaneous retrieval is a bottom-up retrieval process, driven by the high associative strength between the cue and the intended action. Spontaneous retrieval is therefore thought to be reflexive, relatively automatic, and reliable, so long as a strong cue is fully processed in the focus of attention. Critically, multiprocess theory contends that both spontaneous retrieval and monitoring are legitimate routes to successful PM performance. When task interference is present, participants are assumed to be monitoring; by contrast, when task interference is absent (and PM detection is excellent) they are assumed to have spontaneously retrieved the intention.

A critical determinant of whether spontaneous retrieval can occur, or monitoring is required, is the *focality* of the PM task (McDaniel & Einstein, 2000, 2007; Einstein et al., 2005). Though focality can probably be thought of as being on a spectrum, tasks are typically dichotomized as either focal or nonfocal, referring to the degree of processing overlap between the OT and PM task. In cases of high overlap (focal), OT processing simultaneously serves to aid PM cue detection, thereby making retrieval highly likely and task interference unnecessary; in cases of low overlap (nonfocal), by contrast, the OT does not aid PM detection, thus requiring additional PM processing and producing task interference.

To give a concrete example, imagine being given a 2-back task (with word stimuli) as the OT, and the PM task is to press the *Q* key upon seeing the word *history*. Because OT processing overlaps highly with PM task processing—that is, accessing the semantic content of the word is required for both the OT and PM task—the task is focal. In contrast, if the PM task is to press the *Q* key if the word contains the consecutive letters *tor*, the PM task would be considered nonfocal. This is because determining whether the word is the same one as the one presented 2 trials previously does not require processing the individual letters, which the PM task does (see Einstein & McDaniel, 2005, for additional explanation). To reiterate, when given a focal PM task, participants do not have to monitor (though some people may choose to) and typically obtain high PM and little-to-no task interference. Alternatively, when using a nonfocal task, participants must employ additional resources to monitor for the cue, thereby reducing the likelihood of detecting the cue (lower PM) and resulting in task interference.

Multiprocess theory has come to take a dominant stance in the field after the late 2000s (McDaniel & Einstein, 2007), with numerous studies validating the existence of spontaneous retrieval, and reaffirming the qualities and characteristics of monitoring (Abney, McBride, & Petrella, 2013; Boywitt & Rummel, 2012; Brewer, Knight, Marsh, & Unsworth, 2010; Cohen, 2013; Cohen, Jaudas, & Gollwitzer, 2008; Einstein et al., 2005; Harrison & Einstein, 2010; Harrison, Mullet, Whiffen, Ousterhout, & Einstein, 2014; Rummel, Smeekens, & Kane, 2017; Scullin et al., 2010b; Scullin et al., 2010a; Scullin, McDaniel, & Shelton, 2013). Yet, more recently, PM decision control (PMDC) theory (Strickland, Loft, Remington, & Heathcote, 2018), which has been developed considerably from previous instantiations (Heathcote, Loft, &

Remington, 2015; Strickland, Heathcote, Remington, & Loft, 2017), provides an alternative explanation for task interference: a strategic delay.

2.2 Decision Control

According to PMDC theory, participants are assumed to intentionally delay their OT responding, as opposed to monitoring for PM targets, in order to allow additional time (i.e., task interference) for participants to realize the stimulus is a target. The theoretical architecture Strickland et al. (2018) proposed is that OT and PM information simultaneously accumulate toward their respective thresholds for responding, but separately, and at different rates. PM information is thought to accumulate more slowly than OT information, and participants must have some awareness of this, because they increase their OT decision threshold, requiring more information to make a response. The increase in OT conservatism comes at a cost to response times, but makes it more likely that the slower PM information will hit its own threshold and lead to a successful PM response. Hereafter referred to as a strategic or proactive delay, this is one of the primary mechanisms of PMDC, and has its roots in a previous model that only included the delay component (Heathcote et al., 2015). Specifically, based on the analysis of response time and accuracy using accumulator models (e.g., linear ballistic accumulator model, Brown & Heathcote, 2008; Ratcliff diffusion model, Ratcliff, 1978), Heathcote et al. observed a consistent and sizeable increase in only the decision threshold parameter when PM demands were added to the OT, providing evidence favoring a delay strategy, and serving as the basis for PMDC theory.

To better situate the theory, these accumulator models decompose response times and accuracy into a number of parameters, but three have been theoretically important in PM research: *decision threshold* (*a*), *drift-rate* (*v*), and *nondecision time* (*t0*). As alluded to previously, the decision threshold represents the amount of information required before a decision is executed, and Heathcote et al. (2015) contended that increases in this parameter were observed because participants delayed their responding to allow more time for PM information to reach its own threshold. The drift-rate, by contrast, represents the speed at which the information accrues; Heathcote et al. theorized that *this* parameter should have been the one to increase if task interference reflected a limited-capacity monitoring process (discussed further below). Finally, nondecision time represents the total amount of time on a trial devoted to all processes other than decision-making (e.g., response execution and feature encoding).

Therefore, an increase in the threshold parameter indicates a more conservative speed/accuracy policy, causing accuracy and response times to increase. Increases in the drift-rate reflect both faster *and* more accurate responding, and can therefore be seen as indicative of task difficulty, with lower drift-rates signifying more difficult tasks (Ratcliff & Rouder, 1998). Increases in nondecision time reflect slower responding without affecting the decision process; but, some researchers have theorized that PM target-checking behavior before or after an OT decision has been made could be funneled into this parameter (Anderson, Rummel, & McDaniel, 2018; Horn & Bayen, 2015).

To briefly summarize the accumulator modeling efforts thus far, the field has consistently observed increases in the decision threshold parameter (Anderson et al., 2018; Ball & Aschenbrenner, 2018; Boywitt & Rummel, 2012; Heathcote et al., 2015; Horn, Bayen, & Smith, 2011; 2013; Horn & Bayen, 2015; Rummel, Kuhlmann, & Touron, 2013; Strickland et al., 2017, 2018), but contrary to Heathcote et al., some studies have also observed decreases in the drift-rate (Anderson et al., 2018; Boywitt & Rummel, 2012; Horn et al., 2011, 2013; Rummel et al., 2013) and increases in nondecision time (Anderson et al., 2018; Horn & Bayen, 2015). PM-related decreases in the drift-rate conflicts with PMDC because they suggest that alternative processes, which are associated with an increase in the difficulty of the task, contribute to task interference. The same is true of nondecision time, if there is merit to the argument that cue-monitoring could occur before or after the OT decision process.

Returning to the PMDC model, the proactive delay mechanism is only one of several that work in tandem to make PM detection more likely, and this is what separates PMDC from its delay-only predecessor (Heathcote et al., 2015). PMDC theory is based on a model of both the OT and PM task as two separate diffusion processes, instead of making assumptions about an unmodeled PM accumulation process. Based on their findings, Strickland et al. (2018) also proposed, 1) that people can strategically lower their PM threshold, requiring less information to make a PM response, and 2) that participants have reactive control: on PM trials, as more information accumulates suggesting a PM target is present, the OT accumulation process is inhibited to avoid preempting the PM response. Further, PMDC theorists explicitly specify that the proactive control components (OT and PM threshold changes) are “goal-driven”, caused by a “strategic trade-off” between response times and accuracy, and are clearly conscious and deliberate because “participants slow their ongoing task decision process to give a parallel PM process more time to reach response selection” (Strickland et al., 2018, p. 855). Together, these three mechanisms work to prioritize PM information over OT information, slowing down OT responding both proactively and reactively, as well as proactively speeding up the PM decision process via a lower PM threshold.

One final question to address is how PMDC theory handles focality: Under nonfocal conditions, the PM accumulation rate is presumed to be slower than that of the OT, therefore requiring a delay. Plainly, the discussion of PMDC thus far has been assuming a nonfocal PM task. Regarding focal tasks, PMDC assumes that PM information accumulates as fast, or faster, than OT information (Loft & Remington, 2013; Heathcote et al., 2015). In this case, PM information wins the race against OT information without needing any adjustment to the decision threshold (i.e., no task interference), and participants therefore recognize the stimulus as a PM target before reaching an OT decision.

Chapter 3: Leveraging Factors Associated with Task Interference

Focality has repeatedly been shown to be one of the most powerful factors influencing task interference, but there are numerous others to consider, many of which hold theoretical relevance. First, however, it is important to note that even subtle demand characteristics could impact whether or not participants choose to monitor (Anderson, McDaniel, & Einstein, 2017). For example, if the title or description of the study explicitly mentions or prioritizes the PM task, the researcher spends far more time explaining the PM instructions than the OT instructions, or other features clue the participants as to the focus of the study, then participants may be more inclined to monitor (or delay) than they otherwise would.

For researchers interested in minimizing task interference – to study spontaneous retrieval, for instance – there are a number of other strategies that can be used. McDaniel, Umanath, Einstein, and Waldum (2015) recommend using fewer PM targets, prioritizing or emphasizing OT performance, delaying presentation of the first target until well into the experiment, and of course using a focal PM task. These recommendations have been validated in numerous individual studies, and a recent meta-analysis supports these general conclusions (though, the degree of OT emphasis was not assessed) as well: Task interference is reduced with a lower frequency of PM targets, when the first cue is presented later in the experiment, and dramatically so when a focal cue is used instead of a nonfocal cue (Anderson, Strube, & McDaniel, 2019).

Though both PMDC and multiprocess theories have a ready account of task interference, an important point of divergence between the two theories is what *cognitive experiences* are actually associated with manipulations of task interference. Because the underlying theoretical architectures differ so dramatically, it should be possible to look at factors that cause variation in task interference and ask which theory best accounts for the findings, given their fundamental assumptions. For example, a critical assumption of multiprocess theory is that focal tasks are associated with lower task interference and higher PM performance because they have reduced attentional demands, relative to nonfocal tasks, obviating the need for monitoring. With no need to harness attentional resources for the purpose of monitoring, an obvious conclusion would be that the focal task is "easier" than the nonfocal task. Put the other way, nonfocal PM tasks should simultaneously be associated with increases in task interference and with increases in perceived difficulty, relative to focal tasks. This is not limited to manipulations of focality, either; perceived difficulty should increase and decrease along with task interference for any factor.

For PMDC, by contrast, there should be no association between task interference and difficulty ratings because thresholds track an internal policy about how much information you need to reach a decision. Drift-rates, instead, are thought to assess the relative difficulty of a task (Ratcliff & Rouder, 1988), and PMDC clearly states that drift-rates are not affected by PM demands: "nine PM cost datasets…revealed that PM cost is largely attributable to increases in response threshold, and is not attributable to changes in ongoing task evidence accumulation" (Strickland et al., 2018, p. 854). Therefore, any other candidate dependent measure capturing changes in task interference should be voluntary and premeditated, with time being the only real cost. This is because participants set their decision threshold before a trial's stimulus appears, and seemingly without feedback, because the threshold parameter is thought to stabilize within the first few trials of an experiment (Ratcliff, Smith, Brown, & McKoon, 2016). Thus, the cognitive experiences associated with changes in task interference should be both preemptive and intentional, rather than arising because of an inability to control attention. We lack any propositions from PMDC regarding why people may follow different delay policies, or what would cause someone to change their delay policy; therefore, we offer that perhaps the people who are most motivated to perform well on the PM task should display the greatest increases in task interference. It seems plausible that participants who feel the most motivated would be willing to sacrifice the time necessary to ensure detection of the PM targets, whereas the least motivated may simply be eager to finish the experiment quickly, without much concern over missing targets.

To our knowledge there is no published work that has simply asked participants about their perceived difficulty in order to observe whether or not it increases due to possessing a PM intention, but there are several tangential findings that point toward PM-related processes being resource-demanding and costly. First, and perhaps most straightforwardly, successful (nonfocal) PM has been linked to individual differences in working memory. For example, both Smith (2003) and Brewer et al. (2010) showed that those higher in working memory capacity had better PM, and that task interference was associated with PM performance for those with worse working memory. Working memory has been consistently linked to attentionally demanding activities and has been described as being critical for “controlled processing in attention-demanding circumstances” and inhibition of more dominant, automatic processes (Barrett, Tugade, & Engle, 2004, p. 554). Further, given the theoretical import of accumulator model parameters for the current study, research has linked higher working memory to lower drift-rates (i.e., faster information accumulation leading to better task performance) (Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). Thus, the relationship between working memory demands and difficulty appears firm.

Second, we have already noted that several studies have found PM-related reductions in OT drift-rates (Anderson et al., 2018; Boywitt & Rummel, 2012; Horn et al., 2011, 2013; Rummel et al., 2013), offering some evidence that the OT is more difficult when a nonfocal PM intention is also active. Third, the bulk of the literature shows that older adults have worse PM, particularly for nonfocal tasks (Kliegel, Jäger, & Phillips, 2008; Uttl, 2008; 2011), with the theoretical rationale being that older adults have fewer cognitive resources (e.g., working memory capacity) to harness when monitoring for the PM cue (Park, Lautenschlager, Hedden, Davidson, Smith, & Smith, 2002). Finally, nonfocal PM has been consistently associated with sustained activation in fronto-parietal regions such as the anterior prefrontal cortex (Cona, Bisiacchi, Sartori, & Scarpazza, 2016; McDaniel, LaMontagne, Beck, Scullin, & Braver, 2013), which have *also* been linked to working memory processes. Activation in these regions has been described as costly and difficult to maintain for long periods of time, aligning well with the view that these processes are difficult (Braver, Gray, & Burgess, 2007). Thus, in opposition to PMDC theory, there is a strong theoretical basis to expect that task interference should be positively related to difficulty.

Self-reported motivation is less diagnostic than difficulty, because according to multiprocess theory participants may very well be motivated to perform well on the PM task, but this would likely be reflected in their willingness to monitor. Therefore, if they are in fact monitoring, then difficulty ratings should also increase, which would be problematic for PMDC. If, however, motivation ratings were uniquely related to task interference, with no concurrent increase in difficulty, then PMDC would gain strong support. Motivation has been studied in the

PM literature primarily by emphasizing the importance of one task (ongoing or PM), and secondarily by providing some sort of incentive (e.g., money) for good PM performance (see Walter & Meier, 2014, for a review). There has been some variability, but findings generally indicate greater costs when the PM task is emphasized, and that these costs result in better PM (Ball & Brewer, 2018; Ball & Aschenbrenner, 2017; Einstein et al., 2005; Harrison & Einstein, 2010; Horn & Bayen, 2015; Loft & Humphreys, 2012; Loft & Yeo, 2007; Smith & Bayen, 2004). This appears to be particularly true for nonfocal tasks, with focal tasks sometimes showing no effects of task emphasis (Harrison & Einstein, 2010; Kliegel, Martin, McDaniel, & Einstein, 2004). Motivating participants by rewarding successful PM, too, tends to improve performance (Jeong & Cranney, 2009; Krishnan & Shapiro, 1999; Cook, Rummel, & Dummel, 2015), but there is very little research examining the effects of rewards on task interference. That being said, one study – Cook et al. – found no increase in task interference to accompany the PM benefit. On the whole, in support of a PMDC view, there is good evidence that both PM accuracy and task interference are positively related to motivation.

Chapter 4: Aim of the Present Study

To address this theoretical dispute, in the present study we first develop and validate a within-block probe procedure that measures self-reported difficulty and motivation, in order to determine their relationship with task interference. The PMDC model clearly excludes difficulty as having any association with task interference, but one's motivation to perform well on the PM task might. By contrast, monitoring could involve motivational factors—in that some people may not care enough to monitor—but clearly implicates difficulty as the fundamental correlate of task interference. As developed further below, we probed participants throughout the first experiment about the difficulty of the task (both ongoing and PM), as well as their motivation to perform well on the task. Following successful validation of the probe procedure, we applied the methodology in two experiments designed to manipulate task interference, with the central question being, "Is task interference (i.e., change in behavior due to PM) associated with an increase in self-reported difficulty, or is it better tracked by self-reported motivation to perform the PM task?"

Chapter 5: Experiment 1

The purpose of Experiment 1 was to validate the probe methodology as being an accurate assessment of difficulty and motivation, hopefully without affecting behaviors typically associated with performing the OT or the PM task. Critically, we must also assess whether the probes are tapping into *both* the OT and PM task—it is a distinct possibility that participants could base their judgments solely on the OT. Toward that end, we manipulated the difficulty of both the OT and the PM task separately, as well as motivation toward the PM task. To increase difficulty of the OT, we had participants engage in blocks of both 1-back and 2-back trials, with 2-back trials being more difficult (for a review supporting this claim, see Jaeggi, Buschkuehl, Perrig, & Meier, 2010). Thus, we should see increased ratings of difficulty in 2-back blocks relative to 1-back blocks. To increase the difficulty of the PM task, some participants were given one cue for their nonfocal task (the consecutive letters *tor*), whereas others were given three cues for their nonfocal task (the consecutive letters *tor*, *can*, or *mis*)—with three cues being more difficult (Anderson et al., 2019). Therefore, we should see higher difficulty ratings in the three-cue condition than the one-cue condition. Finally, we manipulated motivation toward the PM task by providing a financial incentive dependent on PM task performance in some blocks but not in others. In the motivated blocks we should see greater ratings of motivation than in the non-motivated blocks, given that various motivators, including money, typically improve PM, thereby implying increased recruitment of PM detection processes (Walter & Meier, 2014).

As mentioned previously, it is also important to determine whether the presence of the probes themselves change participants' typical OT or PM task behavior. To assess this, we included a no-probe control group to be compared to probed participants in terms of both OT and

PM performance. Although it is possible the probes will serve as reminders to participants, based on research from Reese and Cherry (2002) that showed no differences in PM accuracy between probed and non-probed groups, we expect no differences in our own study as well.

As an aside, because the probe methodology we develop uses self-report, experimenters should rightfully be wary about their use in novel laboratory paradigms. In PM research, on-line probes have been used with some success to tackle a number of questions not answerable with more objective behavioral measures. As just alluded to, Reese and Cherry (2002) used an open-ended thought probe procedure to assess whether or not on-task and off-task thoughts were associated with PM performance. Although they found no significant relationship between thoughts about the PM task and performance, they did find that PM performance did not differ between probe and no-probe groups. Using a similar methodology, Rummel et al. (2017) showed that participants reported fewer off-task thoughts when they possessed a PM intention as compared to when they did not (implying a more on-task focus), as well as some suggestive findings that PM-related thoughts were associated with better PM performance. They also found that on-task thoughts had no relationship with OT response times, implying that this on-task focus did not manifest itself in terms of increased monitoring behavior. Finally, Anderson and Einstein (2016) successfully used thought probe methodology to illustrate that persisting activation, slowed responding to previously relevant PM targets (i.e., a previously completed PM task), also manifested itself in the form of conscious thoughts about the now-irrelevant PM task. There is therefore some precedent for the use of on-line probes in laboratory PM experiments.

Briefly, we piloted our two less-studied manipulations (motivation and cue-number) before Experiment 1, with 28 participants in the motivation pilot and 27 participants in the cue-number pilot. In both pilots we manipulated the variables within-subjects (counterbalanced). For example, in one block we motivated participants with a financial incentive, and in the other we did not. In the other pilot study, participants had a one-cue PM task (*tor*) for one block and a three-cue PM task (*tor*, *can*, or *mis*) for the other block. In both cases the probe ratings responded sensibly to the manipulation: Participants gave significantly higher motivation ratings in motivated (M = 8.93, SE = .22) relative to non-motivated blocks (M = 7.13, SE = .43), and participants gave significantly higher difficulty ratings in three-cue (M = 6.06, SE = .35) relative to one-cue blocks (M = 4.95, SE = .42). Therefore, we had some confidence in our manipulations before collecting data for Experiment 1.

5.1 Method

5.1.1 Participants and Design

We used a 2 x 2 x 2 mixed factorial design including the within-subjects variables OT load (1-back, 2-back) and motivation (motivated, non-motivated), as well as the between-subjects variable cue number (one, three). We also included a dangling control group that received no probes, thereby assessing the potential impact of the probes on OT and PM task performance. This no-probe control group had a reduced number of participants in each cell of the mixed factorial design described above, for a total of 35 participants (one-cue, N = 20; three-cue, N = 15)[1]. By contrast, we collected data from 87 probed participants (one-cue, N = 43; three-cue, N = 44). Sample sizes were based on a power analysis targeting .80 power for a medium-sized, between-subjects main effect (i.e., of cue number) for the factorial design described above, assuming a within-subject correlation of .50. We obtained .83 power to detect the effect with our final sample size.

[1] Our design was not well-balanced for the no-probe participants because, by chance, we had to eliminate 5 participants in the three-cue condition due computer malfunction or inability to follow instructions.

Participants were Washington University in St. Louis undergraduates who received course credit for compensation, in addition to the monetary compensation earned depending on number of PM targets detected (described in the Procedure). No exclusion criteria were used—any student could participate.

5.1.2 Procedure

Participants were tested in individual rooms and sessions lasted approximately 30 min. They first received instructions for performing the OT, called an *n*-back task, for four blocks of trials. They were told that for two of these blocks they were to determine as quickly as possible whether the presented word was the same as the trial presented immediately prior (1-back). For the other two blocks, they were told to decide whether or not the presented word matched the word presented 2 trials prior (2-back). For match trials, they pressed the "1" key on the number pad; for non-match trials they pressed the "2" key.[2] They then performed 25 practice trials for either the 1-back or 2-back task with response time and accuracy feedback. Of note, when the *n*-back task switched, participants were again given 25 practice trials with the new task.

Following the OT practice, participants were given instructions for the first block, depending on their counterbalancing condition (i.e., motivated 1-back, motivated 2-back, non-motivated 1-back, non-motivated 2-back). During each block, participants were instructed that they were to perform the *n*-back task for this block of trials, but that we had an additional interest in their ability to perform an intention in the future. Specifically, if at any point during the block they saw the consecutive letters *tor* (or in the three-cue condition, *tor*, *can*, or *mis*) they were to

[2] In some rare cases, participants reversed the response keys (i.e., responded 2 for match trials and 1 for non-match trials). This was apparent when participants were nearly perfectly wrong (OT accuracy approaching 0). Because chance accuracy is .50, they clearly mixed up the response keys. For cases obtaining .20 accuracy or worse, in all three experiments, we reverse-scored OT accuracy. There were only two participants across all 3 experiments (both in Experiment 1) who had accuracy near .50, and these people were eliminated from all analyses.

press the *Q* key instead of making their OT decision. They were also told that if the OT response accidentally preempted their PM task response that they could still press the *Q* key a trial or two later. Each of the four blocks had 3 unique PM targets, regardless of cue number condition. Participants were asked to repeat the OT and PM instructions to the experimenter, as an understanding-check, before continuing with each block.

Motivation toward the PM task was manipulated with a PM performance-dependent financial incentive. For the motivated blocks, participants were told that they could earn up to $5 compensation, in addition to the course credit they were entitled to for participation, and that the exact amount of money earned would be proportional to the number of PM targets they detected (but they were not told how many targets would appear). If participants transitioned from a motivated block to a non-motivated block, they were required to tell the experimenter that the financial motivation was over, to ensure understanding. However, if participants started with a non-motivated block, then no mention of any financial incentive was made at that time.

Therefore, the exact counterbalancing procedures used in this experiment was critical. Motivation was always grouped consecutively, to avoid confusion, with participants either executing 2 motivated blocks followed by 2 non-motivated blocks, or vice versa. Similarly, 1-back and 2-back task order was alternated between blocks, starting with a 1-back or 2-back task. For example, half of the participants performed the 1-back followed by the 2-back task for their motivated blocks, and again performed the 1-back followed by the 2-back task for their non-motivated blocks. As follows, there were only 4 counterbalancing possibilities.

Next, participants were given instructions for completing the probes (with the exception of the no-probe control group). They were told that they would be interrupted occasionally during each block of trials and asked a couple questions. First, they were shown the difficulty probe, “On a scale of 1-10, how difficult was it to perform both the [1, 2]-back task and the *Q*-key task together, immediately prior to this prompt? 10 means you were genuinely struggling to perform the tasks, and 1 means it was no challenge at all to perform the tasks.” They were then shown the motivation probe, “On a scale of 1-10, how motivated were you to perform well on both the [1, 2]-back task and the *Q*-key task together, immediately prior to this prompt? 10 means you wanted to achieve your best performance on the tasks, and 1 means you did not care about achieving good performance on the tasks.”[3] Participants were asked to repeat these instructions to the experimenter, as an understanding-check, before continuing. Probe order was maintained to minimize participant confusion, and difficulty was chosen to be the first probe because it is the most theoretically diagnostic.[4]

Participants then proceeded with the first block, which contained 107 trials: 100 OT trials (half match, half non-match), 3 PM trials (always a non-match trial), and 4 probe trials. The PM targets occurred in fixed positions on trials 40, 70, and 105. The probes were also fixed, but on trials 25, 50, 75, and 100, to maximize the temporal distance between the probes. Immediately following, participants were given OT and PM instructions for the next block, according to counterbalancing condition. Task characteristics were nearly the same in the other blocks, but PM targets varied slightly (e.g., on trial 104 instead of 105) to accommodate the *n*-back task randomization.

5.1.3 Materials

Stimuli for the *n*-back task (all words) were selected using the Balota et al. (2007) norms. Items generated were between 4-8 characters in length and had an average Log_Freq_HAL of 8.00. In

[3] A 10-point scale was used because increased Likert scale ranges have been shown to better approach normality than those with reduced ranges, without affecting the general structure of the means and variances (Leung, 2011).
[4] We wanted to allow participants to get into a rhythm of performing the tasks and answering the probes, in an effort to minimize potential error caused by participants’ mixing up the probe questions.

each block, there were 100 OT trials—20 of these were unique words that occurred only once, 20 words appeared on match trials and thus occurred twice, and 20 words appeared on non-match trials but also occurred twice. Each block contained a different set of stimuli.

5.2 Results

5.2.1 OT and PM Performance

We analyzed each dependent variable (OT response times and accuracy, and PM accuracy), using a 2 x 2 x 2 mixed analysis of variance (ANOVA) to assess the impacts of cue number (one, three), OT load (1-back, 2-back), and motivation (low, high). For OT response times and accuracy, we eliminated probe trials, PM trials, and the 3 trials following a PM target from analysis. We trimmed response times below 200 ms or greater than 3 *SD* above the mean (within blocks, between participants). All descriptive statistics are reported in the original metric, but analyses of response times were transformed as closely as possible to normality using the Box-Cox power transformation (for this experiment, response times were raised to the power = .39).[5] Additionally, we eliminated 2 participants who were at near-chance OT accuracy from all analyses. Though OT and PM performance were not the primary dependent variables of interest, we report full analyses to check whether participants behaved in a manner expected by previous PM research.

[5] Despite being uncommon in the field, we transformed response times in Experiment 1 to be analytically consistent across experiments; for the models in Experiments 2 and 3, normality is a critical assumption, so we decided to transform. However, none of the statistical inferences changed for this experiment when analyzed using the raw scores.

Table 1.

Descriptive statistics for Experiment 1.

Number of Cues	Block	Response Time (ms)	Accuracy	Difficulty	Motivation	PM Accuracy
1	M1N1	791 (327)	.96 (.08)	4.20 (2.05)	7.81 (1.98)	.61 (.37)
	M1N2	929 (407)	.96 (.05)	5.76 (2.05)	7.72 (2.07)	.53 (.39)
	M2N1	659 (221)	.97 (.05)	4.22 (2.07)	6.66 (2.22)	.46 (.40)
	M2N2	804 (335)	.95 (.06)	5.55 (1.97)	6.51 (2.35)	.34 (.39)
3	M1N1	960 (377)	.98 (.03)	5.30 (2.20)	7.48 (2.20)	.48 (.37)
	M1N2	1199 (493)	.94 (.12)	6.56 (1.83)	7.31 (2.17)	.50 (.37)
	M2N1	764 (337)	.98 (.04)	5.05 (2.10)	6.31 (2.19)	.34 (.33)
	M2N2	997 (421)	.95 (.05)	6.31 (1.85)	5.96 (2.32)	.29 (.31)

Note. Means are reported with *SD* in parentheses. M1 refers to the presence of monetary motivation, whereas M2 refers to its absence. N1 refers to the 1-back task, whereas N2 refers to the 2-back task. Therefore, for example, M1N2 refers to the motivated 2-back block. OT statistics (response time, accuracy, and PM accuracy) are comprised of $N = 63$ in the one-cue condition and $N = 59$ in the three-cue condition, whereas difficulty and motivation are comprised of $N = 43$ in the one-cue condition and $N = 44$ in the three-cue condition.

The 2 x 2 x 2 ANOVA analyzing response times found a significant main effect of motivation, $F(1,120) = 74.87$, $p < .001$, $MSE = 1.58$, $\eta_p^2 = .38$, indicating that monetarily motivated blocks were slower than the non-motivated blocks (see Table 1 for descriptive statistics). There was also a main effect of OT load, $F(1,120) = 37.90$, $p < .001$, $MSE = 4.02$, $\eta_p^2 = .24$, such that 2-back blocks were slower than 1-back blocks. The final main effect of cue number was also significant, $F(1,120) = 13.19$, $p < .001$, $MSE = 10.77$, $\eta_p^2 = .10$, with participants in the three-cue condition responding more slowly than those in the one-cue condition. There were no significant interactions, all p's > .13. Counterbalancing order was included in a separate model and had a number of significant interactions, which are described in the Appendix, but inclusion of this factor did not change any of the aforementioned inferential conclusions.

OT accuracy had large ceiling effects ($M = .96$, $SD = .07$), preventing any meaningful inferential analyses. Regarding PM accuracy, the only significant effect was the main effect of motivation, $F(1,120) = 38.29$, $p < .001$, $MSE = .096$, $\eta_p^2 = .24$, indicating that participants detected more targets when they were motivated as compared to when they were not. The effect of cue number was marginally significant, $F(1,120) = 3.44$, $p = .07$, $MSE = .252$, $\eta_p^2 = .03$, in the expected direction: participants detected marginally more cues in the one-cue condition than the three-cue condition. A marginal effect was also found for OT load, with numerically greater PM accuracy in the 1-back blocks relative to the 2-back blocks, $F(1,120) = 2.83$, $p = .10$, $MSE = .135$, $\eta_p^2 = .02$. No other effects were significant.

5.2.2 Probe Responses

Of critical interest to this experiment is the validity of the probe responses; therefore, a series of targeted *t*-tests were used. A paired-samples *t*-test comparing motivated to non-motivated blocks was significant, $t(86) = 7.15$, $p < .001$, Cohen's $d = .77$, with higher reported motivation in motivated blocks ($M = 7.58$, $SE = .21$) than non-motivated blocks ($M = 6.36$, $SE = .23$). A second paired-samples *t*-test comparing difficulty ratings for 1-back blocks to 2-back blocks was significant, $t(86) = 10.57$, $p < .001$, Cohen's $d = 1.13$, indicating that reported difficulty was greater for the 2-back task ($M = 6.04$, $SE = .19$) than the 1-back task ($M = 4.70$, $SE = .21$). Finally, an independent-samples *t*-test comparing difficulty ratings between the one-cue and three-cue conditions was significant, $t(85) = 2.38$, $p = .02$, Cohen's $d = .51$, with greater difficulty ratings in the three-cue condition ($M = 5.80$, $SE = .27$) than the one-cue condition ($M =$

4.92, *SE* = .25). Therefore, the probe methodology was validated as assessing what it was designed to assess.[6,7]

Finally, of additional interest was whether the simple act of answering the probes would affect OT or PM task performance. Therefore, an additional 2 independent-samples Wald *t*-tests compared response times and PM accuracy (OT accuracy omitted due to high ceiling effects) — targeting the difference between the 35 no-probe control participants and 87 probed participants. There was no difference between the two groups in response times, $t(71) = 1.58$, $p = .12$, Cohen's $d = .31$. However, probed participants ($M = .48$, $SE = .03$) were significantly more likely than non-probed participants ($M = .34$, $SE = .04$) to detect the PM targets, $t(72.87) = 3.01$, $p = .003$, Cohen's $d = .34$. Likely, the probes served to remind participants about the PM task.

5.3 Discussion

The goal of Experiment 1 was to validate the probe methodology as properly assessing participant's experienced difficulty and motivation toward both the OT and PM task. Examining OT and PM performance, first, we observed significantly slower response times in motivated blocks relative to non-motivated, diverging from prior research (Cook et al., 2015). We observed, as expected by prior work, faster response times for the 1-back blocks than the 2-back blocks (West & Bowry, 2005), and faster response times in the one-cue condition than the three-cue condition (Cohen et al., 2008). Regarding PM accuracy, motivated blocks performed significantly better than non-motivated blocks, converging with Cook et al., and adding

[6] The findings were identical using a more comprehensive ANOVA model, with no interactions by condition. The only finding of additional interest was that participants were significantly ($p = .05$) more motivated when performing the 1-back than the 2-back task.

[7] An interesting point to note is that there was very little variability in difficulty or motivation *within*-blocks. That is, there did not appear to be any time-course effects such as steady changes (e.g., difficulty increase) or large jumps in the probe dependent variables. Instead, variability in probe responses can largely be attributed to either condition or participant differences.

confidence as to the efficacy of our manipulation. The effect of cue-number, however, was only marginally significant, rather than showing a significant increase in PM performance for the one-cue relative to the three-cue condition (see Anderson et al., 2019). The effect of OT load also did not quite reach significance, but this is unsurprising given that previous findings have been inconsistent, with some studies showing that OT load impairs PM performance (West, Krompinger, & Bowry, 2005) and others obtaining no difference (West & Bowry, 2005).

Collectively, these findings align reasonably well with predictions from either of the two theoretical frameworks, PMDC or multiprocess, and conform with the bulk of the literature. Although PMDC theory would not necessarily predict increased response times with three PM cues relative to one—given that both OT and PM information extraction is assumed to operate in parallel, and there is no clear reason why more cues should slow down the PM accumulation rate—increasing the perceived demands of the PM task could encourage participants to allocate a more conservative OT decision threshold. All other results, however, would be straightforwardly predicted by either theory.

Next, confident that participants were behaving in a manner expected by prior research, we move to the primary variables of interest: difficulty and motivation ratings. The probes were validated as measuring what they were intended to measure, with difficulty ratings increasing both when PM and when OT difficulty was increased (i.e., three PM cues compared to one, and 2-back compared to 1-back task, respectively). Additionally, motivation ratings increased when participants were financially incentivized to detect PM targets. We were also interested in determining whether or not the probes themselves affected behavior: the only observed difference was that probed participants had better PM accuracy.

Therefore, an important qualification to be mindful of is that the probe procedure likely reminds participants of the PM task, thereby increasing target detection. Although the effectiveness of reminders in improving PM performance has been somewhat mixed (Guynn, McDaniel, & Einstein, 1998), there is a substantial probability that reminders do improve PM (Henry, Rendell, Phillips, Dunlop, & Kliegel, 2012), and this increased PM accuracy in the probed group. Yet, we do not believe this fundamentally alters the PM processes of interest, for two reasons. First, if reminders increased PM, presumably this would be caused by stimulating participants' monitoring or delay behavior, thereby making target detection more likely. Because that is already the behavior we are interested in, the primary worry, then, would be bringing participants' PM up to ceiling or reducing the likelihood that some participants would *completely* forget about the PM task. Second, we argue that idiosyncratic reminders are likely present in many PM experiments. Often, stimuli are presented with meaning, especially in the commonly used lexical decision task, which involves deciding whether a string of letters forms a real word. In these cases, partial-cuing has been shown to increase PM performance (e.g., seeing *lion* before the actual cue, *tiger*; Taylor, Marsh, Hicks, & Hancock, 2004) and such cues are likely to be idiosyncratically present in many PM experiments (e.g., upon seeing the stimulus *sister*, someone might be reminded of the PM cue *dancer*, if they have a sister who dances). Because the probes are systematically present for everyone, however, participants in the following studies are likely to be reminded of the PM task more, on average, than is typical in PM research. With these qualifications in mind, we proceed to Experiment 2.

Chapter 6: Experiment 2

In Experiment 2, our primary aim was to use the probe methodology to examine the relationship between difficulty, motivation, and task interference. One way to gain leverage on the theoretical debate is to induce changes in task interference over time, and look to see if difficulty or motivation ratings follow a similar pattern. Both PMDC and multiprocess theories assume that participants make an initial judgment about how best to perform the OT and PM task together, and begin the experiment engaging in whatever processes cause task interference (delay or monitoring), if any are needed. Yet, their explanations for why task interference may, in this case, decline over time differs considerably: According to multiprocess, participants may cease monitoring because the heavy reliance on attentional resources makes it prohibitive over long periods of time. From a PMDC perspective, because task interference is due to a preemptive and strategic decision that does not tax attentional resources, motivational factors are likely at play: Perhaps participants wish to get out of the experiment quicker and decide they are willing to miss PM targets should they appear.

We mentioned earlier that manipulating the placement of the first PM target is one way to affect task interference, such that participants are more inclined to monitor/delay if the first target is presented early in the experiment. The speed of encountering the first PM target is referred to as the PM *onset delay*, and the finding that decreases in task interference are associated with increases in the PM onset delay implies that participants' monitoring or delay behavior is affected by their expectancy of target events and/or attentional lapses. Several experiments have manipulated the onset delay, either by waiting a certain amount of time or a certain number of trials to present the first PM target. Conte and McBride (2018), for instance, waited between one and six minutes to present the first PM cue, and found that longer delays lead to less task interference directly before the eventual presentation of the cue, and worse PM performance. In a similar study, McBride et al. (2011) manipulated the onset delay (40, 100, 200, 300, or 400 trials before presentation) and focality. The focal condition was not affected by onset delay; however, at short delays nonfocal PM performance was as good as focal, but then immediately plummeted for all longer delays. Further, task interference was high in the nonfocal condition at the shortest delay, but tapered off for the longer delays, and wasn't present for the focal task except at the shortest delay. Finally, it has also been shown that task interference declines across the experiment more quickly when expected cues are not presented than when they are (Loft, Kearney, & Remington, 2008). Collectively, these findings all point toward the crucial role initial target presentation has in reinforcing monitoring/delay behavior, particularly for nonfocal tasks. If many trials (e.g, 100 or more trials, see McBride et al., 2011) are passed between PM target presentations, then participants either have difficulty maintaining (multiprocess), or choose to reduce (PMDC), high levels of task interference.

In this experiment, we manipulated the onset delay by presenting the first and only PM target either early or late in the experiment. In the *PM-far* condition, the PM target was presented near the end of the experiment, whereas in a *PM-near* condition the target was presented early. We expect task interference to begin at an elevated level, and to generally decline in both conditions, but especially so in the PM-far condition. The reasoning, according to multiprocess theory, is that because monitoring is attentionally demanding, and participants in the PM-far conditions are never reinforced for this behavior, the intention is likely to involuntarily slip from mind as the experiment progresses. Thus, because both conditions go for long periods of time without seeing a PM target, they should both decline, but we expect this reduction in task interference to occur more quickly in the PM-far conditions. Finally, we manipulated PM focality between-subjects, and, following multiprocess theory (McDaniel & Einstein, 2007), we expect effects of task interference to be especially pronounced in the nonfocal conditions. In the focal conditions these effects should be smaller, if present (Einstein et al., 2005).

Of special interest is whether difficulty or motivation ratings will follow a similar pattern to task interference. From the multiprocess perspective, we expect the nonfocal conditions to show an immediate increase in difficulty ratings for the PM block relative to the control block (reduced for focal conditions). Further, because monitoring is difficult, and this behavior is not reinforced in the PM-far conditions, task interference should decline and thus difficulty ratings should as well. In the PM-near conditions, this decline should be less steep because detecting a PM target event should theoretically serve to validate the increased effort. Regarding motivation, predictions are less clear; however, there is a distinct possibility that motivation (i.e., motivation to monitor) could decline when PM targets are not encountered. Yet, it is equally plausible that participants are just as motivated to perform well on the task but cannot maintain monitoring behavior over the course of the entire block. Alternatively, therefore, motivation ratings may not track task interference.

Under PMDC, because setting decision thresholds is a strategic decision not dependent on limited-capacity resources, there is no reason why difficulty ratings should change due to these experimental manipulations. Thus, despite obtaining task interference, there should be no increase in difficulty ratings – in any condition – from the control block to the PM block. Rated difficulty could change over the time-course of the block, but these changes should be identical in the control block and the PM block. Instead, greater task interference should be associated with higher ratings of motivation. The logic is as follows: Under PMDC, participants are presumably aware that nonfocal PM information accumulates slower than the OT information, and if they want to ensure detection of the PM targets then they must raise their decision threshold. Because threshold setting has nothing to do with difficulty, the only reason participants should do this is because they are motivated to perform well. Thus, motivation ratings in the nonfocal conditions should increase in the PM block relative to the control block. Further, if task interference declines in the PM-far conditions as expected, then this should be reflected by decreased motivation, implying that participants decided to respond more liberally, resulting in a lower decision threshold.

As a final note, to aid interpretation of these numerous interactive effects, we denote changes between blocks (PM block – control block) as *PM-related* changes, similar to the way task interference is used to denote a change in response times. Thus, for example, greater PM-related difficulty would mean a larger difference in difficulty ratings between the PM and control blocks, and the same applies for PM-related motivation.

6.1 Method

6.1.1 Participants and Design

The design of this experiment was a 2 x 2 x 2 mixed factorial including the within-subjects variable block (control, PM) and the between-subjects variables focality (focal, nonfocal), and onset delay (PM-near, PM-far). All effects of critical interest involve the interaction with block, therefore, the power analysis targeted a within/between interaction for this factorial design. Obtaining .80 power for a medium-sized, within/between interaction required 10 participants in each of the 4 groups. To increase our likelihood of detecting significant associations (e.g., between dependent measures, such as response times and difficulty ratings), we instead assigned

32 participants to both of the focal groups and 31 participants to both of the nonfocal groups (total $N = 126$), which resulted in .99 power for the within/between interaction. Participants were Washington University in St. Louis undergraduates who received course credit for compensation. As in Experiment 1, no exclusion criteria were used.

6.1.2 Procedure

Participants were tested in individual rooms with sessions lasting approximately 30 min. They first received instructions for performing the OT, a 2-back task, with identical instructions to Experiment 1, and identical practice procedures.

Next, participants were given instructions for completing the first block, depending on counterbalance condition (control, PM), as well as instructions for responding to the probes. As in the first experiment, participants were required to repeat these instructions to the experimenter as an understanding-check. For the control block, participants then completed 204 total trials, with 4 probes embedded on trials 50, 100, 150, and 200. Probes in the control block were different only in that they did not reference the *Q* key task (because they did not possess an intention). For the PM block, PM characteristics changed depending on focality: Half of the participants in the focal conditions received the target *writer*, whereas the other half received the target *dancer*. Half of the participants in the nonfocal conditions received words containing *tor* as their PM target, whereas the other half received words containing *ten*. All PM instructions were identical in form to Experiment 1. The PM block contained 205 total trials. Probes again occurred on trials 50, 100, 150, and 200; in terms of PM targets, either a focal target (*writer*, *dancer*) or a nonfocal target (*tor* word, *ten* word) occurred on trial 25 in the PM-near conditions and on trial 198 in the PM-far conditions.

6.1.3 Materials

As in Experiment 1, stimuli for the *n*-back task (all words) were selected using the Balota et al. (2007) norms. Items generated were between 4-8 characters in length and had an average Log_Freq_HAL of 8.00. In each block, there were 200 OT trials—40 of these were unique words that occurred only once, 40 words appeared on match trials and thus occurred twice, and 40 words appeared on non-match trials but also occurred twice. Each block contained a different set of stimuli.

6.2 Results

6.2.1 Analysis Strategy

We used hierarchical linear modeling to analyze clusters of trials (level 1) nested within blocks (level 2) nested within participants (level 3), letting intercepts vary. Predicting OT response times, for example, we computed means within clusters of trials split in time by the probe trials. As in Experiment 1, we eliminated probe trials, PM trials, and the 3 trials following a PM target from in our OT response time and accuracy analyses. We also used the same trimming procedures and optimally transformed the data using the Box-Cox power transformation (response times raised to the power = .43).[8] Unfortunately, OT accuracy was again on ceiling in both blocks (control, M = .97, SD = .03; PM, M = .96, SD = .03), preventing meaningful inferential analysis. Regarding PM performance, because there was only 1 PM target, we used logistic regression to examine group differences in the factorial design.

[8] As in Experiment 1, using raw response times in the following model, instead of transformed, did not affect any inferential conclusions.

In the hierarchical model, we predicted each continuous dependent variable by block (control, PM), focality (focal, nonfocal), and onset delay (PM-near, PM-far), all dummy coded (0, 1) respectively. Further, we included another factor, *quarter*, by splitting each block into clusters of trials reflecting 4 time points (Q1, Q2, Q3, Q4). Finally, we included counterbalance condition as an individual factor. To arrive at the best possible model, we compared three nested models for each dependent variable – main effects and all two-way interactions, all three-way interactions, and the four-way interaction – using a likelihood ratio test.[9] The best model for all three dependent measures included only two-way interactions between the four factors.

We first report these univariate analyses, and then assess the simultaneous contributions of the four dependent variables in separating the groups using MANOVA (multivariate ANOVA). The use of MANOVA is particularly informative in this study because we are interested in the relations among dependent variables (i.e., task interference, PM-related difficulty and motivation ratings, and PM accuracy), and their differential patterning among the conditions. The optimal combination of the dependent variables allows us to examine their relative weights to assess how well each variable captures group variance, while simultaneously accounting for their covariance structure: it could be the case that difficulty and motivation ratings explain the same group variance, for instance. Descriptive statistics and the correlations among variables can be found in Tables 2 and 3, respectively, and the full results from each model can be found in Tables 4 (response times), 5 (difficulty), and 6 (motivation), but we focus our report on theoretically informative findings (i.e., involving an interaction with block).

[9] We began with a model including two-way interactions, rather than main effects only, because interactions with the variable *block* are critically informative to assess any PM-related effects.

Table 2.

Descriptive statistics for Experiment 2.

Focality	Onset Delay	Block	Response Time (ms)	Accuracy	Difficulty	Motivation	PM Accuracy
Focal	PM-Near	Control	541 (145)	.97 (.03)	3.84 (1.56)	6.94 (1.92)	
		PM	592 (176)	.96 (.04)	4.23 (2.11)	7.06 (1.73)	.84 (.37)
	PM-Far	Control	632 (195)	.98 (.02)	3.79 (1.73)	7.76 (1.85)	
		PM	646 (202)	.96 (.04)	4.45 (1.70)	7.63 (1.76)	.56 (.50)
Nonfocal	PM-Near	Control	609 (154)	.97 (.03)	4.27 (1.92)	7.70 (1.94)	
		PM	803 (241)	.96 (.03)	5.02 (2.25)	7.64 (1.95)	.71 (.46)
	PM-Far	Control	647 (188)	.98 (.02)	3.98 (1.52)	7.15 (2.10)	
		PM	841 (293)	.96 (.02)	5.26 (2.24)	7.58 (1.96)	.39 (.50)

Note. Means are reported with *SD* in parentheses. Both focal conditions contain $N = 32$, and both nonfocal conditions contain $N = 31$.

Table 3.

Correlation table of the variables in Experiment 2.

	Focality	Onset Delay	Quarter	Block	Response Time	Difficulty	Motivation	PM Accuracy
Focality	1							
Onset Delay	0	1						
Quarter	0	0	1					
Block	0	0	0	1				
Response Time	.27	.12	-.19	.24	1			
Difficulty	.13	.01	-.004	.17	.003	1		
Motivation	.04	.05	-.04	.02	.14	-.10	1	
PM Accuracy	-.16	-.31	0	0	.13	-.02	-.07	1

Note. Focality = focal, nonfocal (coded 0, 1); Onset Delay = PM-near, PM-far (coded 0, 1); Quarter = ranges from 1-4, reflecting the segment of each block; Block = control, PM (coded 0, 1); Response Time = transformed from ms; Difficulty = ranges from 1-10; Motivation = ranges from 1-10; PM Accuracy = miss, hit (coded 0, 1).

6.2.2 Response Times

Examining response times first, there was a significant interaction between focality and block, *b* = 1.48, *t* = 5.56, *p* < .001, reflecting greater slowing in the PM block relative to the control block (i.e., task interference) for nonfocal compared to focal conditions (see Figure 1). As seen in Figure 2, there was also a significant interaction between quarter and block, *b* = -.35, *t* = 2.65, *p* = .008, indicating a reduction in task interference between quarters one and four. As an aside, for the reader interested in spontaneous retrieval, a paired-samples *t*-test indicated that overall task interference was marginally significant in the focal conditions (*M* slowing = 33 ms, *SD* = 138), *t*(63) = 1.96, *p* = .06, Cohen's *d* = .24.

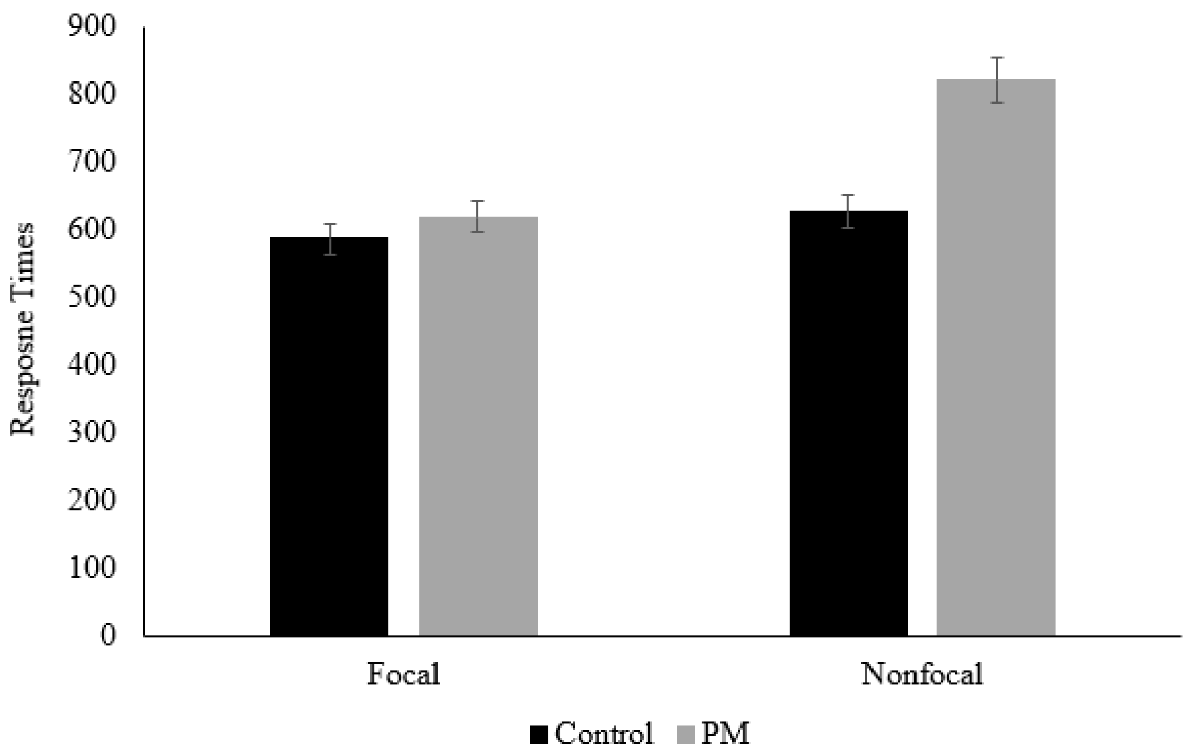

Figure 1. Depicts the significant interaction (Experiment 2) for response times between block (control, PM) and focality (focal, nonfocal). Values have been converted back to the original metric (ms) from their transformation. Error bars represent standard error of the mean.

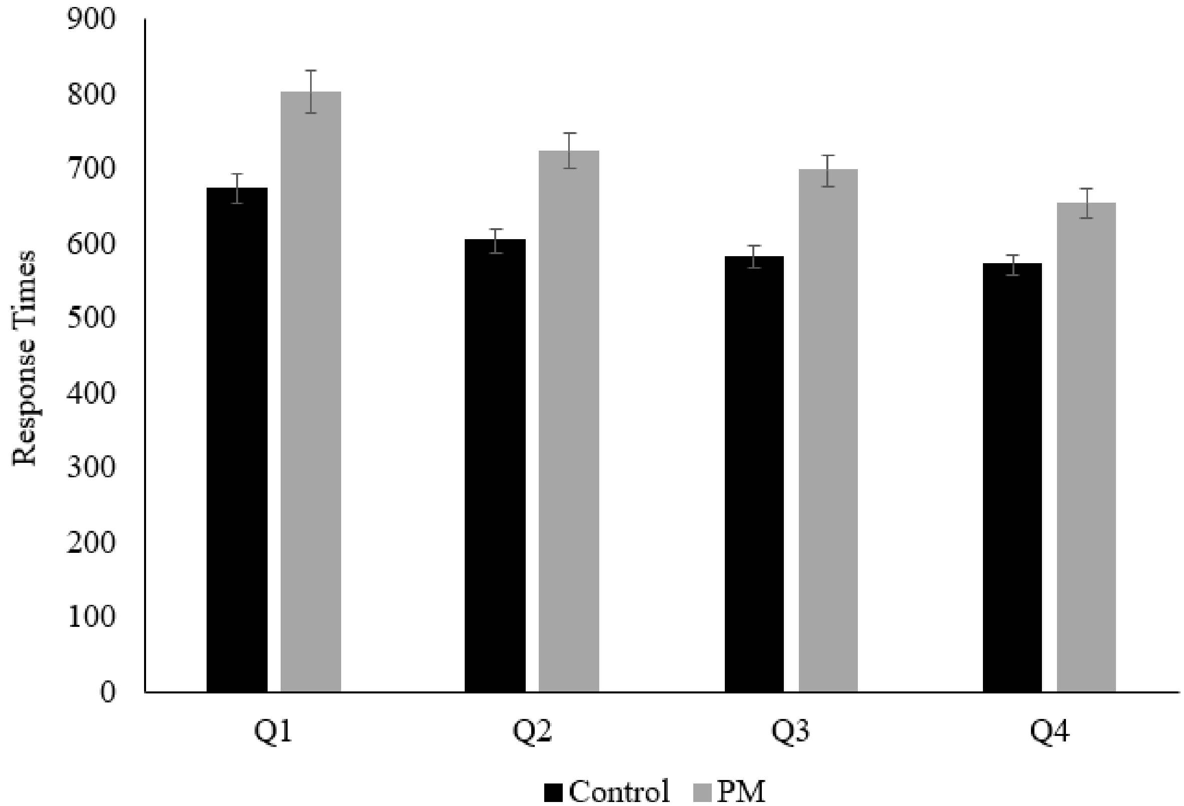

Figure 2. Depicts the significant interaction (Experiment 2) for response times between block (control, PM) and quarter (Q1, Q2, Q3, Q4). Values have been converted back to the original metric (ms) from their transformation. Error bars represent standard error of the mean.

Table 4.

Experiment 2 model predicting response times.

	Response Time			
Predictors	*b*	*SE*	*t*	*p*
(Intercept)	**15.06**	**.38**	**40.05**	**< .001**
Block	**.58**	**.24**	**2.40**	**.02**
Focality	.93	.47	1.98	.05
Onset Delay	.91	.47	1.94	.06
Quarter 2 (ref 1)	**-.60**	**.13**	**-4.60**	**< .001**
Quarter 3 (ref 1)	**-.72**	**.13**	**-5.49**	**< .001**
Quarter 4 (ref 1)	**-.79**	**.13**	**-6.01**	**< .001**
Counterbalance	.22	.31	.68	.50
Block:Focality	**1.48**	**.27**	**5.56**	**< .001**
Block:Onset Delay	-.26	.27	-.97	.33
Block:Quarter 2 (ref 1)	-.03	.13	-.19	.85
Block:Quarter 3 (ref 1)	-.03	.13	-.21	.83
Block:Quarter 4 (ref 1)	**-.35**	**.13**	**-2.65**	**.008**
Focality:Onset Delay	-.44	.63	-.69	.49
Focality:Quarter 2 (ref 1)	-.17	.13	-1.30	.19
Focality:Quarter 3 (ref 1)	**-.32**	**.13**	**-2.45**	**.01**
Focality:Quarter 4 (ref 1)	**-.46**	**.13**	**-3.47**	**< .001**
Onset Delay:Quarter 2 (ref 1)	.008	.13	.06	.95
Onset Delay:Quarter 3 (ref 1)	-.06	.13	-.43	.67
Onset Delay:Quarter 4 (ref 1)	-.02	.13	-.14	.89
Model Characteristics				
$ICC_{Subject}$			0.63	
ICC_{Block}			0.24	
Deviance			2974	
Observations			1008	

Note. Response times were transformed by raising to the power = .43. Significant effects are noted in bold. Block = control, PM (coded 0, 1); Focality = focal, nonfocal (coded 0, 1); Onset Delay = PM-near, PM-far (coded 0, 1); Quarter = ranges from 1-4, reflecting the segment of each block. Counterbalance = control block first, PM block first (coded 0, 1).

6.2.3 Difficulty and Motivation Ratings

Overall, participants reported being fairly motivated (*M* = 7.44, *SD* = 1.91) and did not find the task to be overly difficult (*M* = 4.35, *SD* = 1.90). Next, we analyzed participants' ratings of difficulty with the same model used to analyze OT response times. In this model (see Table 5), there was a marginally significant interaction between focality and block, $b = .49$, $t = 1.83$, $p = .07$, suggesting a greater increase in PM-related difficulty for nonfocal relative to focal conditions (see Figure 3). Next, the interaction between block and quarter was significant, with quarters two, $b = -.56$, $t = 2.49$, $p = .01$, and three, $b = -.40$, $t = 1.78$, $p = .08$, both showing a decline in PM-related difficulty relative to quarter one (Figure 4). When predicting motivation ratings, the best model lacked any interactions with block; therefore, having no relationship with task interference, reported motivation was not theoretically informative in this experiment.

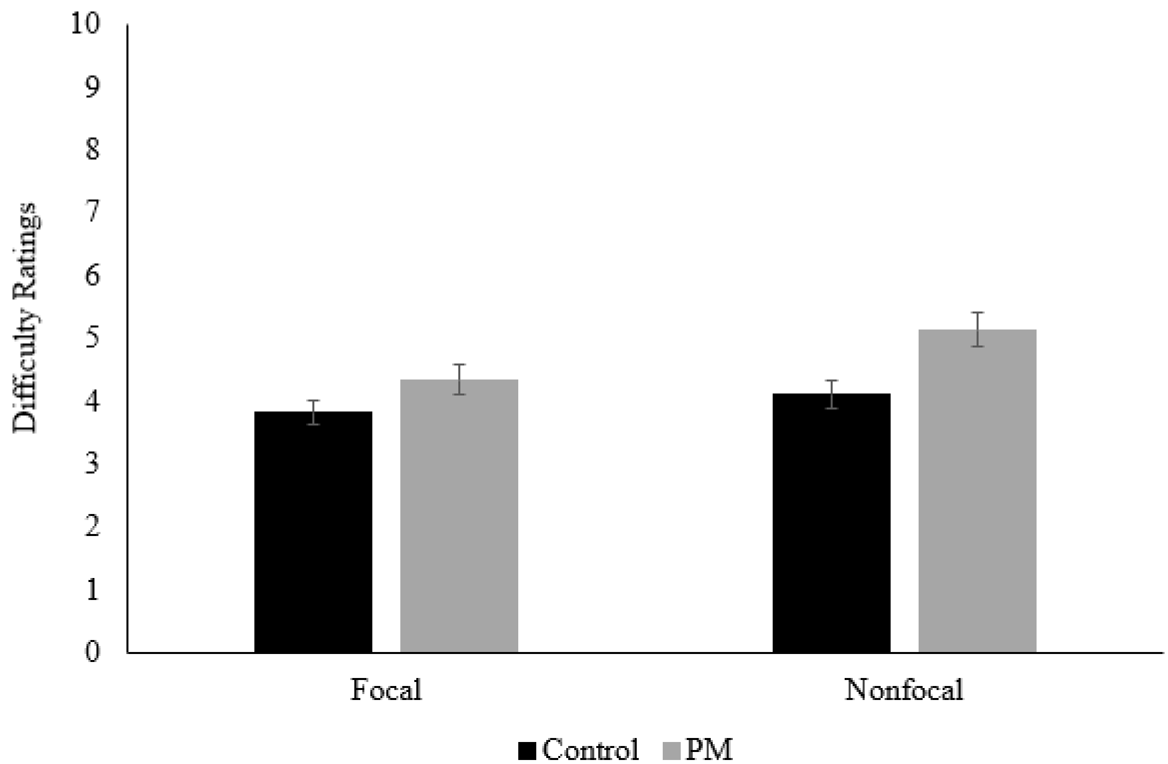

Figure 3. Depicts the marginally significant interaction (Experiment 2) for difficulty ratings between block (control, PM) and focality (focal, nonfocal). Error bars represent standard error of the mean.

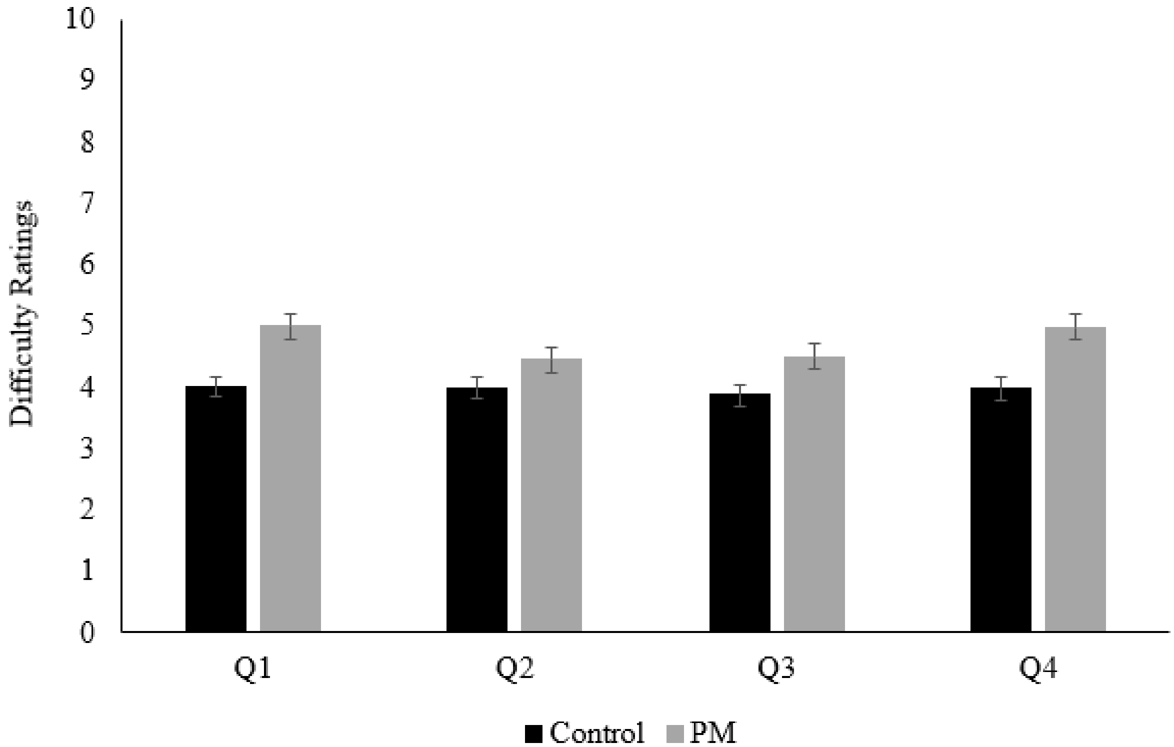

Figure 4. Depicts the significant interaction (Experiment 2) for difficulty ratings between block (control, PM) and quarter (Q1, Q2, Q3, Q4). Error bars represent standard error of the mean.

Table 5.

Experiment 2 model predicting difficulty ratings.

			Difficulty	
Predictors	*b*	*SE*	*t*	*p*
(Intercept)	**4.04**	**.38**	**10.51**	**< .001**
Block	**.59**	**.27**	**2.21**	**.03**
Focality	.54	.47	1.15	.25
Onset Delay	-.08	.47	-.17	.86
Quarter 2 (ref 1)	.08	.22	.37	.71
Quarter 3 (ref 1)	.15	.22	.66	.51
Quarter 4 (ref 1)	-.01	.22	-.04	.97
Counterbalance	-.49	.31	1.83	.12
Block:Focality	.49	.27	1.83	.07
Block:Onset Delay	.38	.27	1.44	.15
Block:Quarter 2 (ref 1)	**-.56**	**.22**	**-2.49**	**.01**
Block:Quarter 3 (ref 1)	-.40	.22	-1.78	.08
Block:Quarter 4 (ref 1)	-.02	.22	-.07	.94
Focality:Onset Delay	-.13	.61	-.21	.84
Focality:Quarter 2 (ref 1)	-.20	.22	-.89	.37
Focality:Quarter 3 (ref 1)	-.15	.22	-.68	.50
Focality:Quarter 4 (ref 1)	-.26	.22	-1.14	.26
Onset Delay:Quarter 2 (ref 1)	.05	.22	.21	.83
Onset Delay:Quarter 3 (ref 1)	-.35	.22	-1.56	.12
Onset Delay:Quarter 4 (ref 1)	.25	.22	1.14	.26
Model Characteristics				
$ICC_{Subject}$			0.51	
ICC_{Block}			0.15	
Deviance			3768	
Observations			1008	

Note. Significant effects are noted in bold. Block = control, PM (coded 0, 1); Focality = focal, nonfocal (coded 0, 1); Onset Delay = PM-near, PM-far (coded 0, 1); Quarter = ranges from 1-4, reflecting the segment of each block. Counterbalance = control block first, PM block first (coded 0, 1).

Table 6.

Experiment 2 model predicting motivation ratings.

	Motivation			
Predictors	*b*	*SE*	*t*	*p*
(Intercept)	**7.33**	**.38**	**19.12**	**< .001**
Block	-.15	.20	-.76	.45
Focality	.51	.48	1.06	.29
Onset Delay	.61	.48	1.28	.20
Quarter 2 (ref 1)	-.03	.17	-.19	.85
Quarter 3 (ref 1)	-.13	.17	-.75	.45
Quarter 4 (ref 1)	**-.57**	**.17**	**-3.41**	**< .001**
Counterbalance	-.29	.32	-.90	.37
Block:Focality	.17	.20	.85	.40
Block:Onset Delay	.10	.20	.52	.61
Block:Quarter 2 (ref 1)	.04	.17	.24	.81
Block:Quarter 3 (ref 1)	.17	.17	1.00	.32
Block:Quarter 4 (ref 1)	.25	.17	1.52	.13
Focality:Onset Delay	-1.02	.65	-1.57	.12
Focality:Quarter 2 (ref 1)	.17	.17	1.02	.31
Focality:Quarter 3 (ref 1)	.02	.17	.12	.90
Focality:Quarter 4 (ref 1)	.20	.17	1.20	.23
Onset Delay:Quarter 2 (ref 1)	.06	.17	.33	.74
Onset Delay:Quarter 3 (ref 1)	-.17	.17	-1.00	.32
Onset Delay:Quarter 4 (ref 1)	.32	.17	1.90	.06
Model Characteristics				
$ICC_{Subject}$			0.70	
ICC_{Block}			0.09	
Deviance			3271	
Observations			1008	

Note. Significant effects are noted in bold. Block = control, PM (coded 0, 1); Focality = focal, nonfocal (coded 0, 1); Onset Delay = PM-near, PM-far (coded 0, 1); Quarter = ranges from 1-4, reflecting the segment of each block. Counterbalance = control block first, PM block first (coded 0, 1).

6.2.4 PM Accuracy

We used logistic regression to predict the likelihood of detecting the single PM target with focality, onset delay, and their interaction as factors, as well as counterbalancing order. There was a significant effect of onset delay, $b = .71$, $z = 3.46$, $p < .001$, with a greater likelihood of detecting the target in the PM-near conditions ($M = .78$, $SE = .05$) than the PM-far conditions ($M = .48$, $SE = .06$). The effect of focality was only marginal, $b = .38$, $z = 1.85$, $p = .07$, but in the expected direction: there was a marginally greater probability of detecting the PM target in focal conditions ($M = .70$, $SE = .06$) than in nonfocal conditions ($M = .55$, $SE = .06$). The interaction was not significant ($p = .92$).

6.2.5 MANOVA

Finally, we computed difference scores between PM and control blocks for response times, difficulty, and motivation, and used these three dependent variables – as well as PM accuracy – in a 2 x 2 MANOVA with focality and onset delay as independent variables. There was a significant effect for both focality, $F(4, 119) = 8.65$, Wilks' lambda = .78, $p < .001$, and onset delay, $F(4, 119) = 4.33$, Wilks' lambda = .87, $p = .003$, but no interaction between the two ($p = .40$). Dependent variable weights can be found in Table 7, and correlations between variables in Table 8, but generally speaking the nonfocal conditions were separated from the focal conditions primarily by an increase in task interference and PM-related ratings of difficulty, whereas the PM-near conditions were separated from the PM-far conditions primarily by an increase in PM accuracy and task interference, but a decline in PM-related difficulty. In the individual ANOVA models, focality was significantly predictive of both task interference, $F(1, 122) = 30.56$, $p < .001$, $MSE = 2.30$, $\eta_p^2 = .20$, and PM-related difficulty, $F(1, 122) = 3.95$, $p = .05$, $MSE = 2.32$, η_p^2

= .03. Onset delay, by contrast, was significantly predictive of PM accuracy only, $F(1, 122) = 13.53$, $p < .001$, $MSE = .21$, $\eta_p^2 = .10$.

Table 7.

MANOVA dependent variable weights for Experiments 2 and 3.

Experiment	Independent Variable	Task Interference	PM-related Difficulty	PM-related Motivation	PM Accuracy
2	**Focality**	**-.75**	**-.27**	**-.09**	**.08**
	Onset Delay	**.13**	**-.22**	**-.06**	**.15**
	Focality*Onset Delay	.08	.04	.19	-.01
3	**Focality**	**-1.12**	**-.32**	**.20**	**.15**
	Emphasis	-.20	-.08	-.07	.01
	Focality*Emphasis	-.02	-.08	-.17	.01

Note. Significant effects are noted in bold. Task interference is the transformed change in response times from control to PM block, and PM-related difficulty and motivation are similarly difference scores between blocks. Focality reflects the change from nonfocal to focal, Onset delay changes from PM-far to PM-near, and Emphasis changes from PM-emphasis to OT-emphasis. PM accuracy in Experiment 2 only had one target, whereas in Experiment 3 there were four.

Table 8.

Correlation table of the variables in the Experiment 2 MANOVA.

	Focality	Onset Delay	Task Interference	PM-related Difficulty	PM-related Motivation	PM Accuracy
Focality	1					
Onset Delay	0	1				
Task Interference	**.45**	-.08	1			
PM-related Difficulty	**.18**	.14	**.21**	1		
PM-related Motivation	.08	.05	.16	-.11	1	
PM Accuracy	-.16	**-.31**	-.03	-.15	-.007	1

Note. Significant effects are noted in bold. Focality = focal, nonfocal (coded 0, 1); Onset Delay = PM-near, PM-far (coded 0, 1); Task Interference = transformed from ms; PM-related Difficulty = ranges from 1-10; PM-related Motivation = ranges from 1-10; PM Accuracy = miss, hit (coded 0, 1).

6.3 Discussion

In accordance with the bulk of the prior literature, we predicted greater task interference for the nonfocal conditions relative to the focal conditions, greater task interference in the PM-near conditions than in the PM-far conditions, and an interaction between the two: the effect of onset delay would be larger for the nonfocal conditions. Put another way, we expected PM effects on response times to be reduced, if present, for the focal conditions. We also predicted that task interference would begin at an elevated level (e.g., Q1), but that this difference would be markedly reduced by the end of the block (e.g., Q4). Finally, we thought that the reduction in task interference would interact with onset delay, such that a more pronounced decline would be observed in the PM-far than the PM-near conditions.

However, the best model contained two-way interactions only; therefore, any three-way interactive hypotheses such as, "nonfocal conditions should exhibit greater task interference in the PM-near conditions than the PM-far conditions", were clearly not supported. In addition, we received no evidence for any effect of onset delay on task interference. What we *did* observe was less task interference in focal relative to nonfocal conditions, and that overall task interference was significantly reduced by the end of the block. If the PMDC model is seen as adaptive, in that task interference is allowed to change within a block based on metacognitive expectations or task characteristics (see Anderson et al., 2019 for details), then these findings align well with either multiprocess or PMDC. Similarly, both theories can easily handle the PM accuracy findings: The theories predict that declining task interference should result in lower PM performance when the target is presented later in the experiment (see also, Scullin et al., 2010b). This is because whatever behavior task interference reflects (e.g., monitoring or delay) is assumed to benefit PM performance. Similarly, marginally greater PM performance in focal conditions is theoretically nondiagnostic—according to either theory, high focal PM performance does not require task interference.

In regard to the null findings for onset delay, we suspect that controlling for overall target number between conditions may have eliminated the presence of an effect. This is because we based our prediction on a meta-analytically reliable effect, and meta-analyses do not control all other variables to isolate effects (Anderson et al., 2019). For a fact, the absolute number of targets is greater in experiments using an earlier onset delay. Or, reversed, if the experimenter waits until well into the experiment to present the first target (long onset delay), then there is pragmatically less time to present additional targets. Thus, experiments containing earlier onset delays are more likely to reinforce those participants to monitor/delay after the first PM target by presenting additional targets, thereby increasing task interference relative to later onset delays. There was no opportunity for this to occur in our experiment: both onset delay conditions started out with high task interference, then only the PM-near condition participants (at least, those who were able to detect the target) got reinforced. However, *all* participants subsequently experienced a long gap between targets, likely resulting in identical task interference patterns between conditions.

More importantly, turning to the probe ratings, the multiprocess perspective predicts that difficulty should parallel task interference. Given the task interference findings, therefore, multiprocess theory predicts higher difficulty ratings in PM blocks relative to control blocks (PM-related difficulty), that this should generally decline from Q1 to Q4, and greater PM-related difficulty ratings for nonfocal relative to focal conditions. By contrast, the PMDC model would likely predict that motivation ratings should parallel task interference, but certainly predicts no relationship between difficulty ratings and task interference.

We observed (marginally) greater PM-related difficulty ratings for nonfocal conditions, as well as a decline in PM-related difficulty ratings across quarters (specifically, ratings were reduced in Q2 and Q3 relative to Q1), clearly supporting a multiprocess view. There was a slight bump in PM-related difficulty in Q4; speculatively, this could signify fatigue effects, but it also may be due to some participants failing to detect any PM targets (especially in the PM-far conditions) despite being well into the experiment. This could have caused some participants to believe targets had been presented but missed, and subsequently judged the task to be more difficult than previously imagined.[10] Finally, there was no relationship between block and motivation, implying no PM-related effects on motivation, thereby offering little support for PMDC. These conclusions do not change when accounting for covariance between dependent measures—the MANOVA attributed task interference and PM-related difficulty ratings as the primary measures distinguishing between focal and nonfocal tasks, not motivation.

[10] When conditionalized on whether or not participants detected the target, there were no significant effects, but there was some indication that participants in the PM-far conditions, relative to PM-near conditions, had greater PM-related difficulty in Q4 when they missed the PM target (*M* difference in task interference between the onset delay conditions = 1.77) than when it was detected (*M* difference = 1.29).

Chapter 7: Experiment 3

The purpose of Experiment 3 was to use an instructional manipulation of task interference to once again see whether or not PM-related difficulty or motivation ratings would follow. Toward that end, we assigned participants to either a PM-emphasis condition or an OT-emphasis condition, providing instructions regarding the relative importance of either the OT or PM task. Specifically, participants in the PM-emphasis condition were told to treat the PM task as their primary task, whereas OT-emphasis participants were told the 2-back task was primary. We again manipulated focality between subjects, for a total of four conditions. Based on prior work, we anticipate that nonfocal conditions and PM-emphasis conditions will obtain greater task interference (Smith & Bayen, 2004), with less interference in OT-emphasis conditions and focal conditions (Harrison & Einstein, 2010; Kliegel et al., 2004). The focal conditions actually should not result in any task interference (e.g., Einstein et al., 2005), though participants may choose to monitor when given PM-emphasis instructions (despite it being unnecessary). The PM-emphasis conditions should also have better PM performance than the OT-emphasis conditions. Once again, multiprocess theory predicts that difficulty ratings will align with task interference, whereas PMDC predicts that it should not, and perhaps instead motivation should.

To better detail these hypotheses, from the monitoring viewpoint, we anticipate that in the nonfocal conditions, PM-emphasis instructions will increase difficulty ratings relative to the OT-emphasis instructions. Based on findings from Kliegel et al. (2004), we anticipate that PM or OT emphasis instructions will not have any effect in the focal conditions, with both groups likely obtaining minimal task interference and no PM-related increase in difficulty. Though theoretically uninformative from the multiprocess view, PM-emphasis instructions may or may not increase self-reported motivation: Motivation could increase if participants tend to prioritize the PM task more heavily than the OT, but motivation may also increase equally between the two conditions.

According to PMDC, emphasizing either the OT or PM task should make participants more likely to set a conservative threshold to the ongoing task. However, because PM-emphasis instructions typically increase task interference relative to prioritizing the OT, participants in the PM-emphasis conditions should set even higher thresholds. Assuming task interference is greater when the PM task is prioritized, motivation ratings should track these threshold differences. However, because PMDC theory contends that the focal PM accumulation rate is as fast as the OT accumulation rate, focal PM should be insensitive to motivation, assuming participants choose not to expand their decision threshold. It could be argued, however, that properly motivated participants with a focal PM task (e.g., in the focal PM-emphasis group) may obtain both task interference and an increase in PM-related motivation. In the nonfocal conditions, predictions are more straightforward: The highest motivation ratings should be present in the nonfocal PM-emphasis condition, followed by the nonfocal OT-emphasis condition. Once again, PMDC predicts no changes in difficulty ratings from the control to the PM block, nor any differences in PM-related difficulty ratings among any of the four conditions.

7.1 Method

7.1.1 Participants and Design

In this experiment, we again used a 2 x 2 x 2 mixed factorial design including the within-subjects variable block (control, PM), and the between-subjects variables focality (focal, nonfocal), and task emphasis (OT-emphasis, PM-emphasis). Power analyses were identical to Experiment 2, and we assigned 32 participants to every condition other than the nonfocal OT emphasis condition, which received 30 participants (total N = 126). Participants were Washington University in St. Louis undergraduates who received course credit for compensation. As in the first two experiments, no exclusion criteria were used.

7.1.2 Procedure

Procedures were identical to those used in Experiment 2; however, we changed the between-subjects manipulation of onset delay to one of task emphasis. Participants in the PM-emphasis conditions were told before the PM block, “When performing the 2-back task and the Q-key task, we would like you to focus your efforts PRIMARILY on the Q-KEY TASK. This is your most important goal.” Participants in the OT-emphasis conditions, by contrast, were told before the PM block, “When performing the 2-back task and the Q-key task, we would like you to focus your efforts PRIMARILY on the 2-BACK TASK. This is your most important goal.”

We also included more PM target events in this experiment than in Experiment 2. Out of the 205 trials in the PM block, the 4 PM targets were presented on trials 38, 94, 143, and 198 (with a couple slight variations of one or two trials). The control block was 204 trials again, with probes occurring on trials 50, 100, 150, and 200 for both blocks.

7.1.3 Materials

Stimuli were identical to those used in Experiment 2, other than three non-match trials, which were swapped for PM targets.

7.2 Results

7.2.1 Analysis Strategy

We used an identical analysis plan as was used in Experiment 2. That is, we analyzed clusters of trials (level 1) nested within blocks (level 2) nested within participants (level 3), letting intercepts vary. We also used the same trimming procedures and optimally transformed response times using the Box-Cox power transformation (response times raised to the power = .46).[11] OT accuracy was again on ceiling in both blocks (control, M = .97, SD = .04; PM, M = .97, SD = .03), preventing analysis.

We compared three nested models (main effects only and all two-way interactions, all three-way interactions, and the four-way interaction) for each dependent variable using a likelihood ratio test. Just as in Experiment 2, the best model for all dependent variables included only two-way interactions between the four factors. Regarding PM performance, because there were 4 PM targets, we used ordinary regression to determine whether or not there were group differences. We first report these univariate analyses, and then use the same MANOVA design, this time with focality and task emphasis as independent variables. Descriptive statistics and the correlations among variables can be found in Tables 9 and 10, respectively, and the full results from each model can be found in Tables 11 (response times), 12 (difficulty), and 13 (motivation). Again, we focus our report on theoretically informative findings (i.e., involving an interaction with block).

[11] Once again, transforming response times, instead of using raw scores, did not affect any inferential conclusions in the following model.

Table 9.

Descriptive statistics for Experiment 3.

Focality	Emphasis	Block	Response Time (ms)	Accuracy	Difficulty	Motivation	PM Accuracy
Focal	OT	Control	666 (210)	.97 (.04)	4.18 (1.74)	6.95 (1.92)	
		PM	665 (179)	.97 (.03)	4.72 (1.66)	6.93 (1.91)	.82 (.25)
	PM	Control	563 (145)	.97 (.03)	4.15 (1.56)	6.66 (2.35)	
		PM	592 (123)	.97 (.02)	5.02 (1.80)	7.11 (2.16)	.78 (.30)
Nonfocal	OT	Control	691 (243)	.97 (.04)	4.42 (1.97)	7.10 (2.29)	
		PM	881 (306)	.97 (.03)	5.76 (1.99)	7.02 (2.02)	.49 (.31)
	PM	Control	615 (185)	.97 (.02)	4.38 (1.86)	7.56 (2.06)	
		PM	820 (225)	.96 (.03)	5.73 (1.94)	7.27 (2.07)	.49 (.34)

Note. Means are reported with *SD* in parentheses. The focal PM-emphasis group contains $N = 30$, and all other conditions contain $N = 32$.

Table 10.

Correlation table of the variables in Experiment 3.

	Focality	Emphasis	Quarter	Block	Response Time	Difficulty	Motivation	PM Accuracy
Focality	1							
Emphasis	.02	1						
Quarter	0	0	1					
Block	0	0	0	1				
Response Time	.26	-.16	-.19	.23	1			
Difficulty	.13	.01	.04	.24	.02	1		
Motivation	.07	.04	-.06	.004	-.001	-.01	1	
PM Accuracy	-.46	-.04	0	0	.04	-.20	-.05	1

Note. Focality = focal, nonfocal (coded 0, 1); Emphasis = OT emphasis, PM emphasis (coded 0, 1); Quarter = ranges from 1-4, reflecting the segment of each block; Block = control, PM (coded 0, 1); Response Time = transformed from ms; Difficulty = ranges from 1-10; Motivation = ranges from 1-10; PM Accuracy = miss, hit (coded 0, 1).

7.2.2 Response Times

Examining response times first, there was a significant interaction between focality and block, *b* = 2.21, *t* = 5.82, *p* < .001, replicating the finding of greater task interference for nonfocal compared to focal conditions (see Figure 5). Next, the interaction between quarter and block was significant, *b* = -.56, *t* = 2.93, *p* = .004, once again replicating a reduction in task interference between quarters one and four (Figure 6). As an aside, for the reader interested in spontaneous retrieval, a paired-samples *t*-test indicated that overall task interference was *not* present in the focal conditions (*M* slowing = 14 ms, *SD* = 164), *t*(63) = .92, *p* = .36, Cohen's *d* = .12.

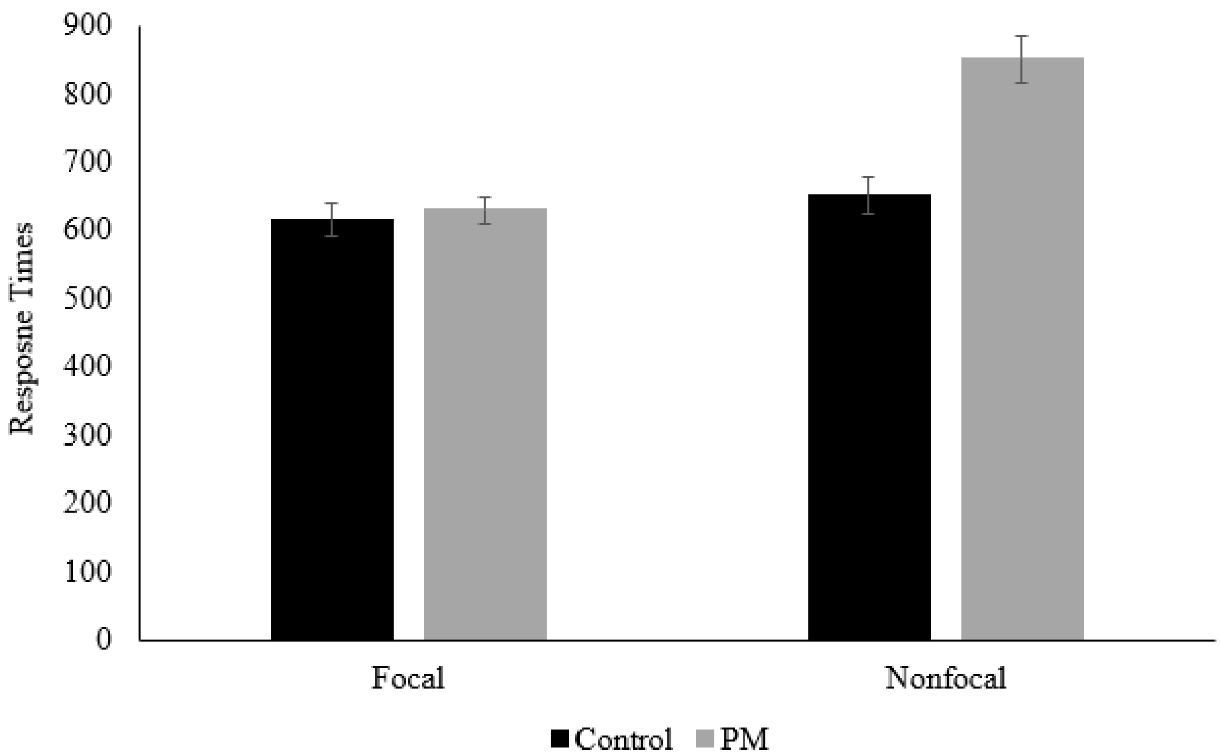

Figure 5. Depicts the significant interaction (Experiment 3) for response times between block (control, PM) and focality (focal, nonfocal). Values have been converted back to the original metric (ms) from their transformation. Error bars represent standard error of the mean.

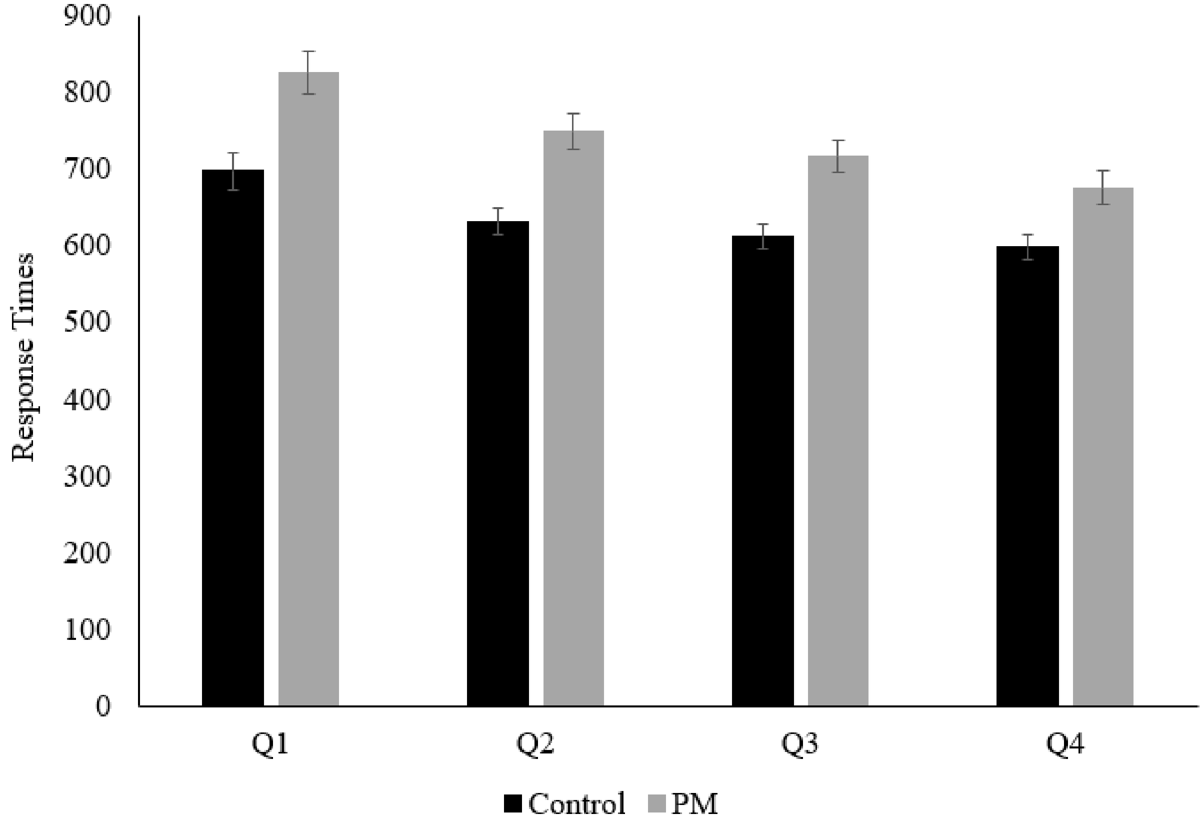

Figure 6. Depicts the significant interaction (Experiment 3) for response times between block (control, PM) and quarter (Q1, Q2, Q3, Q4). Values have been converted back to the original metric (ms) from their transformation. Error bars represent standard error of the mean.

Table 11.

Experiment 3 model predicting response times.

		Response Time		
Predictors	*b*	*SE*	*t*	*p*
(Intercept)	**19.88**	**.49**	**40.36**	**< .001**
Block	.28	.35	.82	.42
Focality	.69	.61	1.14	.26
Emphasis	-1.17	.62	-1.88	.06
Quarter 2 (ref 1)	**-.59**	**.19**	**-3.11**	**.002**
Quarter 3 (ref 1)	**-.65**	**.19**	**-3.45**	**< .001**
Quarter 4 (ref 1)	**-.71**	**.19**	**-3.473**	**< .001**
Counterbalance	-.12	.41	-.29	.78
Block:Focality	**2.21**	**.38**	**5.82**	**< .001**
Block:Emphasis	.39	.38	1.04	.30
Block:Quarter 2 (ref 1)	-.09	.19	-.47	.64
Block:Quarter 3 (ref 1)	-.22	.19	-1.15	.25
Block:Quarter 4 (ref 1)	**-.56**	**.19**	**-2.93**	**.004**
Focality:Emphasis	.36	.81	.45	.66
Focality:Quarter 2 (ref 1)	-.29	.19	-1.54	.12
Focality:Quarter 3 (ref 1)	**-.55**	**.19**	**-2.90**	**.004**
Focality:Quarter 4 (ref 1)	**-.72**	**.19**	**-3.80**	**< .001**
Emphasis:Quarter 2 (ref 1)	-.11	.19	-.60	.56
Emphasis:Quarter 3 (ref 1)	-.26	.19	-1.35	.18
Emphasis:Quarter 4 (ref 1)	**-.41**	**.19**	**-2.13**	**.03**
Model Characteristics				
$ICC_{Subject}$			0.56	
ICC_{Block}			0.28	
Deviance			3684	
Observations			1008	

Note. Response times were transformed by raising to the power = .46. Significant effects are noted in bold. Block = control, PM (coded 0, 1); Focality = focal, nonfocal (coded 0, 1); Emphasis = OT emphasis, PM emphasis (coded 0, 1); Quarter = ranges from 1-4, reflecting the segment of each block. Counterbalance = control block first, PM block first (coded 0, 1).

7.2.3 Difficulty and Motivation Ratings

As in Experiment 2, participants reported being fairly motivated (M = 7.08, SD = 2.09) and did not find the task to be overly difficult (M = 4.80, SD = 1.83). Next, we analyzed difficulty ratings with the same two-way interactions model used to analyze response times. There was a significant interaction between focality and block, b = .64, t = 2.45, p = .02, reflecting greater PM-related difficulty in nonfocal compared to focal conditions (Figure 7). As seen in Figure 8, there were three marginally significant interactions between quarter and block, suggesting that PM-related difficulty declined in quarters two (b = -.39, t = 1.78, p = .08), three (b = -.39, t = 1.76, p = .08), and four (b = -.42, t = 1.88, p = .06) relative to quarter one.

For the same model predicting motivation ratings, there was a marginally significant interaction, b = -.39, t = 1.97, p = .05, whereby focal conditions showed a small increase in PM-related motivation, but nonfocal conditions actually showed a small decline in PM-related motivation. Lastly, there was a marginal interaction between quarter and block, b = .28, t = 1.85, p = .07, suggesting a PM-related increase in motivation ratings in quarter four relative to quarter one.

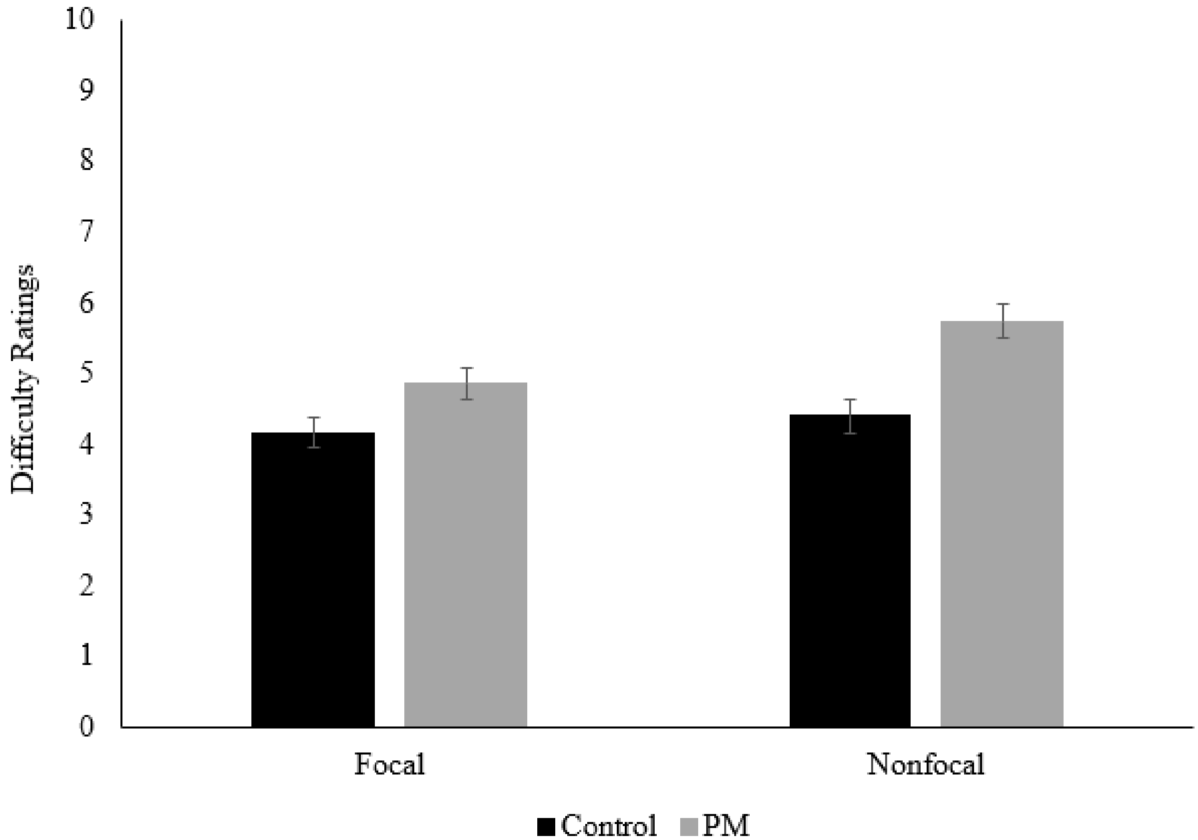

Figure 7. Depicts the significant interaction (Experiment 3) for difficulty ratings between block (control, PM) and focality (focal, nonfocal). Error bars represent standard error of the mean.

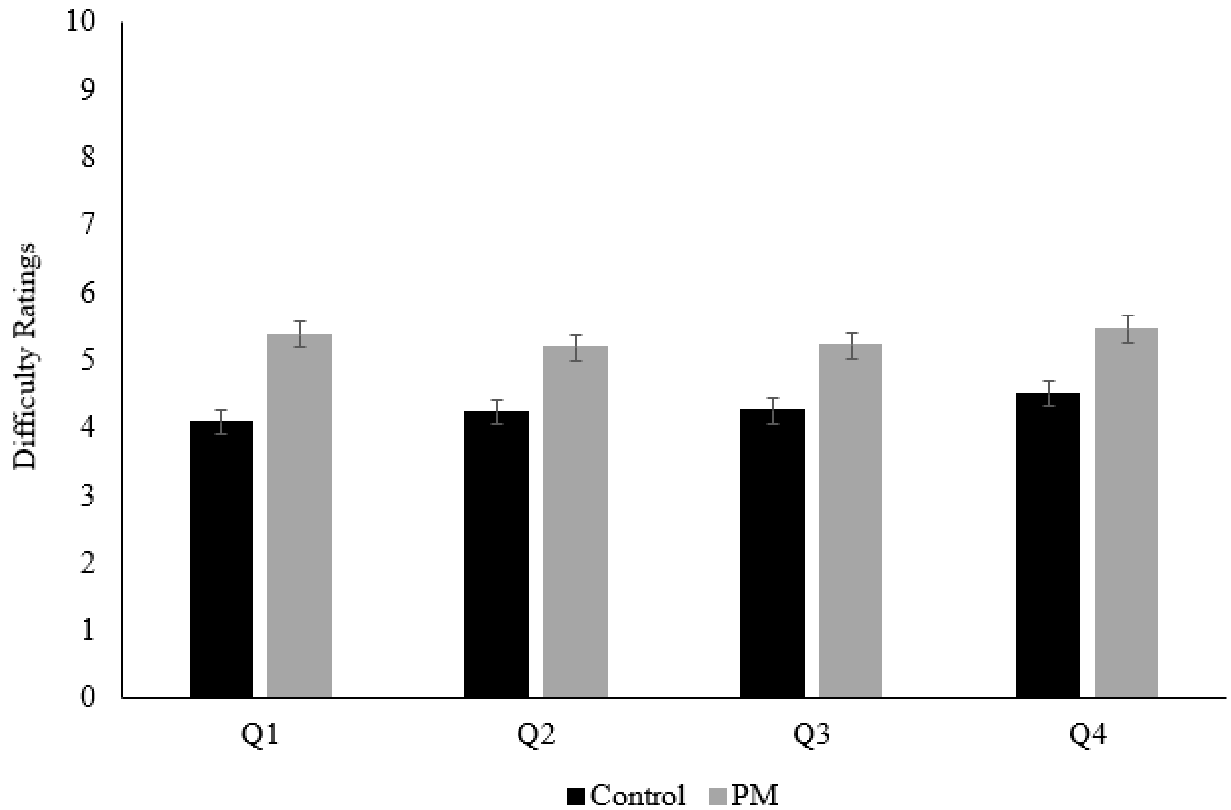

Figure 8. Depicts the marginally significant interaction (Experiment 3) for difficulty ratings between block (control, PM) and quarter (Q1, Q2, Q3, Q4). Error bars represent standard error of the mean.

Table 12.

Experiment 3 model predicting difficulty ratings.

			Difficulty	
Predictors	*b*	*SE*	*t*	*p*
(Intercept)	**3.94**	**.37**	**10.58**	**< .001**
Block	**.92**	**.26**	**3.51**	**< .001**
Focality	.27	.45	.60	.55
Emphasis	.28	.46	.61	.54
Quarter 2 (ref 1)	.30	.22	1.34	.18
Quarter 3 (ref 1)	.38	.22	1.70	.09
Quarter 4 (ref 1)	**.49**	**.22**	**2.22**	**.03**
Counterbalance	-.21	.30	-.71	.48
Block:Focality	**.64**	**.26**	**2.45**	**.02**
Block:Emphasis	.16	.26	.62	.54
Block:Quarter 2 (ref 1)	-.39	.22	-1.78	.08
Block:Quarter 3 (ref 1)	-.39	.22	-1.76	.08
Block:Quarter 4 (ref 1)	-.42	.22	-1.88	.06
Focality: Emphasis	-.17	.59	-.29	.77
Focality:Quarter 2 (ref 1)	-.04	.22	-.17	.86
Focality:Quarter 3 (ref 1)	.06	.22	.26	.80
Focality:Quarter 4 (ref 1)	.18	.22	.84	.40
Emphasis:Quarter 2 (ref 1)	-.23	.22	-1.06	.29
Emphasis:Quarter 3 (ref 1)	-.44	.22	-1.98	.05
Emphasis:Quarter 4 (ref 1)	-.25	.22	-1.14	.26
Model Characteristics				
$ICC_{Subject}$			0.50	
ICC_{Block}			0.16	
Deviance			3708	
Observations			1008	

Note. Significant effects are noted in bold. Block = control, PM (coded 0, 1); Focality = focal, nonfocal (coded 0, 1); Emphasis = OT emphasis, PM emphasis (coded 0, 1); Quarter = ranges from 1-4, reflecting the segment of each block. Counterbalance = control block first, PM block first (coded 0, 1).

Table 13.

Experiment 3 model predicting motivation ratings.

	Motivation			
Predictors	*b*	*SE*	*t*	*p*
(Intercept)	**6.99**	**.42**	**16.76**	**< .001**
Block	.02	.19	.09	.93
Focality	.28	.52	.53	.60
Emphasis	-.14	.53	-.27	.79
Quarter 2 (ref 1)	-.10	.15	-.68	.50
Quarter 3 (ref 1)	-.26	.15	-1.74	.08
Quarter 4 (ref 1)	**-.64**	**.15**	**-4.31**	**< .001**
Counterbalance	.26	.36	.73	.47
Block:Focality	-.40	.20	-1.97	.05
Block:Emphasis	.13	.20	.66	.51
Block:Quarter 2 (ref 1)	.03	.15	.17	.86
Block:Quarter 3 (ref 1)	.20	.15	1.32	.19
Block:Quarter 4 (ref 1)	.28	.15	1.85	.07
Focality:Emphasis	.40	.72	.56	.58
Focality:Quarter 2 (ref 1)	-.007	.15	-.04	.97
Focality:Quarter 3 (ref 1)	-.003	.15	-.02	.98
Focality:Quarter 4 (ref 1)	.15	.15	1.00	.32
Emphasis:Quarter 2 (ref 1)	-.02	.15	-.12	.90
Emphasis:Quarter 3 (ref 1)	.07	.15	.44	.66
Emphasis:Quarter 4 (ref 1)	.05	.15	.36	.72
Model Characteristics				
$ICC_{Subject}$			0.77	
ICC_{Block}			0.09	
Deviance			3110	
Observations			1008	

Note. Significant effects are noted in bold. Block = control, PM (coded 0, 1); Focality = focal, nonfocal (coded 0, 1); Emphasis = OT emphasis, PM emphasis (coded 0, 1); Quarter = ranges from 1-4, reflecting the segment of each block. Counterbalance = control block first, PM block first (coded 0, 1).

7.2.4 PM Accuracy

We used ordinary regression to predict the proportion of PM targets detected with focality, task emphasis, and their interaction as factors, as well as counterbalancing order. There was a significant effect of focality, $b = .15$, $z = 5.73$, $p < .001$, with a greater likelihood of detecting the target in the focal conditions ($M = .80$, $SE = .03$) than the nonfocal conditions ($M = .49$, $SE = .04$). There was also an effect of counterbalancing, $b = -.11$, $z = 2.08$, $p = .04$, such that PM performance was better when the PM block came second ($M = .70$, $SE = .04$) than when it came first ($M = .58$ $SE = .04$). Neither task emphasis, nor its interaction with focality, were significant (both *p*'s > .64).

7.2.5 MANOVA

Finally, the MANOVA model used PM accuracy and difference scores between PM and control blocks for response times, difficulty, and motivation, as dependent variables in a 2 x 2 MANOVA with focality and task emphasis as independent variables. There was a significant effect for focality, $F(4, 119) = 17.52$, Wilks' lambda = .63, $p < .001$, with no effect of task emphasis or an emphasis by focality interaction (both *p*'s > .47). Dependent variable weights are in Table 7, and correlations between variables in Table 14. We replicated the finding from Experiment 2 that nonfocal conditions were separated from the focal conditions primarily by an increase in task interference and greater PM-related difficulty. Likely due to increasing the number of PM targets from Experiment 2, in this experiment lower PM accuracy was also associated with having a nonfocal PM task. The PM-related motivation weighting did not replicate, and actually flipped signs. In the individual ANOVA models, focality significantly predicted task interference, $F(1, 122) = 33.66$, $p < .001$, $MSE = 4.70$, $\eta_p^2 = .21$, PM-related difficulty, $F(1, 122) = 6.08$, $p = .02$, $MSE = 2.16$, $\eta_p^2 = .05$, and PM accuracy, $F(1, 122) = 32.54$, $p < .001$, $MSE = .09$, $\eta_p^2 = .21$.

Table 14.

Correlation table of the variables in the Experiment 3 MANOVA.

	Focality	Task Emphasis	Task Interference	PM-related Difficulty	PM-related Motivation	PM Accuracy
Focality	1					
Task Emphasis	.02	1				
Task Interference	**.46**	.09	1			
PM-related Difficulty	**.22**	.06	**.22**	1		
PM-related Motivation	-.17	.05	-.006	-.08	1	
PM Accuracy	-.11	-.11	**.29**	-.16	-.05	1

Note. Significant effects are noted in bold. Focality = focal, nonfocal (coded 0, 1); Task Emphasis = OT-emphasis, PM-emphasis (coded 0, 1); Task Interference = transformed from ms; PM-related Difficulty = ranges from 1-10; PM-related Motivation = ranges from 1-10; PM Accuracy = proportion detected out of 4 targets.

7.3 Discussion

In this experiment, we assigned participants to either a PM-emphasis condition or an OT-emphasis condition, and they were given either a focal or nonfocal task during the PM block. We expected to observe task interference in nonfocal, but not focal, conditions (Kliegel et al., 2004), and for task interference to be greater when the PM task was emphasized (Smith & Bayen, 2004). However, if there was a small amount of interference for focal tasks, we expected it to be limited to the PM-emphasis condition (Harrison & Einstein, 2010). To reiterate, multiprocess theory predicts that difficulty ratings will align with task interference, whereas PMDC predicts that they should not, thus requiring some other factor to explain differences in task interference (e.g., motivation).

Surprisingly, we found no significant interactions between task emphasis and block for any of the dependent measures; that is, there were no PM-related changes in any dependent variable due to task emphasis for this experiment. Without a standard PM control condition, it is impossible to determine whether or not both emphasis conditions spent more time (i.e., monitoring or delay) on their respective tasks of import. Therefore, it could easily be the case that both conditions increased task interference at an equivalent level; or, it could be the case that neither condition responded to the manipulation as predicted. Yet, the latter interpretation is favored, given that the amount of slowing between blocks was comparable (approximately 200 ms) for nonfocal conditions between Experiments 2 and 3, and Experiment 2 received no such emphasis manipulation. In addition, we observed no PM accuracy benefit in the PM-emphasis relative to OT-emphasis conditions, which is quite surprising. The reasons for why this may have happened are not apparent, given the strength of the instructions, their similarity to prior research (e.g., Harrison & Einstein, 2010; Kliegel et al., 2004; Smith & Bayen, 2004), and the fact that participants were required to repeat the instructions to the experimenter. Another potential reason is this: We know from Experiment 1 that probing participants increases their PM accuracy, so it is also possible that participants in the OT-emphasis condition were able to perform better than expected by being reminded periodically of the PM task. Regardless of the reasons, lacking any effect of task emphasis simply removes one of our variables of influence on task interference.

Focusing on differences we did obtain, we largely replicated findings from Experiment 2: Nonfocal conditions exhibited greater task interference and PM-related difficulty ratings than focal conditions, and both dependent variables declined across quarters. We also replicated the extremely consistent focal task benefit for PM performance (Kliegel et al., 2008). Although there were PM-related changes in motivation ratings (an increase for focal but decrease for nonfocal, and a bump in quarter four relative to quarter one), neither effect follows a predicted pattern according to either theory. For one, there is no reason to expect an increase in experienced motivation in either focal conditions or toward the end of the block (Q4), given that task interference declines in both of those scenarios. Second, there is also no reason to expect a decline in PM-related motivation ratings for nonfocal conditions, given the high levels of task interference observed for nonfocal tasks. The latter is especially problematic for PMDC, given that participants with a nonfocal task would need to be at least as motivated as those with a focal task, or they should not have increased their thresholds. On the other hand, multiprocess theory could account for these results by claiming that extended periods of monitoring may be draining and discouraging, thereby hurting participants' motivation (though this would not necessarily be predicted). Collectively, these results once again support a multiprocess perspective, and the MANOVA findings offer additional support by consistently implicating higher task interference and PM-related difficulty, but not motivation, as separating nonfocal from focal conditions. Interestingly, the weighting of PM-related motivation was increased relative to Experiment 1, but this is likely due to the somewhat odd interaction between focality and block in the hierarchical model (described above).

Chapter 8: General Discussion

The motivating question behind this research was simple: When we observe task interference due to some PM-related behavior, monitoring or delay, what cognitive experiences are associated with this increase in response time? As detailed at length throughout, under the multiprocess view, the assumption is that task interference reflects cognitively demanding cue-monitoring processes; thus, the most likely candidate is an increase in subjective task difficulty. According to PMDC, by contrast, task interference reflects a strategic decision to follow a more conservative OT decision policy – trading speed for accuracy – thereby allowing more time for PM information to accrue. Therefore, we offered motivation as a potential candidate explaining why participants may choose to increase or decrease their proactive delay policies. Regardless, according to PMDC, difficulty ratings should be unrelated to task interference.

To address this question, we first developed and validated a probe procedure asking participants to self-report their on-line experiences of difficulty and motivation (separately) toward the ongoing and PM tasks (jointly). In Experiment 1, we demonstrated that self-reported difficulty increased when the OT was made objectively more difficult (2-back compared to 1-back), as well as when the PM task was made more difficult (three-cue compared to one-cue). We also demonstrated that self-reported motivation increased when there was a strong motivator present (i.e., money). Thus, the probe procedure responded as expected, with subjective ratings reliably tracking objective changes in both difficulty and motivation.

Next, confident in the sensitivity of our measures, we conducted two experiments designed to manipulate task interference in conceptually different ways. In Experiment 2, we tried to induce changes in task interference longitudinally by changing the amount of time (i.e., number of trials) preceding the first PM target, also known as the PM onset delay. In Experiment 3, instead, we varied whether or not the OT or PM task was primary (i.e., task emphasis), in the hopes of pushing task interference around with a more preemptive, global task approach. In both experiments we also manipulated focality, which nearly always results in greater task interference and worse PM accuracy for nonfocal PM tasks, relative to focal tasks (Anderson et al., 2019).

8.1 A Consistent Story

Beginning with Experiment 2, we initially predicted that the PM-near conditions should result in greater task interference, because detecting the target earlier in the experiment should reinforce behaviors associated with PM detection, causing a slower decline in monitoring/delay behavior as each quarter of the experiment progressed. Instead, we did not obtain any PM-related differences between onset delay conditions for any dependent variable—other than PM accuracy, for which the PM-near conditions had a greater likelihood to detect the single PM target than the PM-far conditions. Though it was somewhat surprising we did not obtain any effects of onset delay on task interference, in hindsight the reason why seems clear: if the first target is presented later in the experiment (i.e., PM-far), then there is less time for the experimenter to place additional targets, thus reducing the overall potential for monitoring/delay behavior to be reinforced. This was especially true here, because both conditions received only one PM target. Thus, it is likely that both groups' task interference declined at statistically identical rates because we did not include additional targets in the PM-near conditions.

Though onset delay did not influence task interference as expected, focality did, with far greater slowing in nonfocal conditions than focal conditions. Thus, this study was still able to address the critical question: Given that task interference was greater in nonfocal conditions, did these participants also have a larger PM-related increase in rated difficulty than those with a focal task? Additionally, given that task interference tended to decline across quarters, did PM-related difficulty ratings also decline? The answer to both questions was unambiguously yes, offering strong support for the multiprocess view that task interference indicates monitoring. Motivation ratings did not add any additional wrinkles to the story, either. Even the MANOVA models, after accounting for the covariance between motivation ratings and the other dependent measures, consistently associated greater task interference and difficulty ratings, but not motivation ratings, with nonfocal conditions. It is also worth noting that motivation ratings, despite being high in both experiments, were normally distributed and not on ceiling, so a purely statistical explanation likely cannot account for the null findings.

Turning to Experiment 3, the story was identical: PM task emphasis instructions, relative to OT emphasis, did not result in greater slowing as we expected, but focality once again had a powerful influence, replicating results from Experiment 2. Though it is unclear whether both conditions slowed responding to an equivalent extent, or whether neither condition responded to the task emphasis manipulation, the most important finding was that self-reported difficulty once again tracked task interference and motivation ratings did not. Specifically, PM-related difficulty ratings increased in nonfocal conditions relative to focal, and declined across quarters of the experiment, just as in Experiment 2.

8.2 Theoretical Implications

Since the early 2000s (e.g., Smith, 2003), slowed OT responding has been observed when participants must simultaneously perform an embedded PM task, and numerous studies following have consistently implicated attentionally-demanding cue-monitoring processes in what has been termed task interference (or cost). Yet, despite findings that tangentially implicate difficulty, such as in neuroimaging studies (e.g., sustained activation is usually observed in areas of the fronto-parietal network for nonfocal tasks; McDaniel et al., 2013) or studies examining relations to working memory (e.g., working memory capacity is important for nonfocal PM; Brewer et al., 2010), to our knowledge no study has yet to take the (deceptively) simple approach to directly ask participants how difficult they find the task to be.

With recent theoretical advancements that instead implicate a proactive delay as the process underlying task interference (Heathcote et al., 2015; Strickland et al., 2017, 2018), this question has become more relevant than ever; PMDC rightfully forces the field to stop, back-track, and make sure that we are operating under correct assumptions. To help adjudicate this debate, we examined not only task interference, but also variables that should theoretically be associated with task interference. Although PMDC theory does not explicitly connect what sorts of behavioral changes might be associated with changes in task interference, their proposed mechanism for task interference, a strategic delay, does place some limits on the possibilities. For example, because participants set their decision threshold before onset of the stimulus, and some degree of stability is obtained within the first few trials (Ratcliff et al., 2016), participants must base their delay policy on task characteristics, such as focality (see Strickland et al., 2017). Further, because the drift-rate parameter, and not the threshold parameter, is thought to track difficulty (Ratcliff & Rouder, 1998), and PMDC theorists contend that the drift-rate does not (and should not) decrease when nonfocal PM demands are present, there is no reason to think that difficulty ratings should be related to task interference.

Thus, we offered motivation as a factor that could potentially explain changes in task interference, with the following reasoning: people who are particularly motivated to perform well on the PM task should do whatever they can (within reason) to ensure target detection, and the more motivated they are, the greater lengths they should be willing to go to. Therefore, these highly motivated individuals should follow a very conservative speed/accuracy policy (resulting in high task interference); in contrast, the less motivated participants should not change the base delay policy they adhere to when performing the OT only. If PMDC still has merit, then our intuitions in this regard were obviously off, given that motivation ratings were completely unrelated to PM, in either Experiment 2 or 3, in any sensible way. Also mentioned in previous work (Anderson & McDaniel, 2019; Anderson et al. 2018, 2019), forthcoming advancements from PMDC theorists should work to address this knowledge gap, and inform the field what factors influence people's decisions to change their thresholds.

Furthermore, PMDC theory must now accommodate the findings presented here, which suggest that difficulty is reliably related to task interference. That is, future theoretical work from PMDC theorists must in some way account for the PM-related changes in difficulty ratings we observed, such as a greater increase in experienced difficulty for nonfocal PM tasks. However, even the simple increase in difficulty ratings from control block to PM block is problematic for PMDC, given the theoretical framework that has been proposed (Strickland et al., 2018). No obvious reconciliation presents itself immediately, and Experiment 1 severely limits worries over the validity of the difficulty measure itself. Further, the fact that our findings replicate between Experiments 2 and 3 should ease concerns as to their reliability. Therefore, PMDC faces a difficult theoretical challenge ahead.

PMDC could potentially accommodate these findings by contending that participants' perceptions of difficulty are influenced by their perceptions of task complexity; Strickland et al. (2018) did state that task interference in nonfocal tasks could be due either to deliberately trying to increase PM detection, or it could be due to increased "perceptions of task complexity" (p.

874). If perceived task complexity causes changes in perceived difficulty, without a corresponding change in *objective* difficulty, then PMDC could reconcile some of our findings. However, in that case it would still be uncertain why perceptions of task complexity would decline across quarters of the experiment, as we have shown PM-related difficulty and task interference to decline. Further, if this line of reasoning has merit, then the increase in thresholds should produce a corresponding increase in OT accuracy. Though OT accuracy was on ceiling in this experiment, a meta-analysis of task interference recently found that accuracy does not increase—and instead declines—as task interference increases (Anderson et al., 2019). Thus, there is good evidence that objective task difficulty does increase along with task interference, and the findings presented here provide evidence that participants can reliably report their subjective experiences of this difficulty increase.

Multiprocess theory, on the other hand, has little trouble handling this study's findings. In fact, the multiprocess view straightforwardly predicted nearly every result across all three experiments. The best support comes from the finding that focal tasks obtained a smaller bump in difficulty, from control to PM block, than nonfocal tasks. This is because the explanation for why focal tasks typically result in lower task interference and higher PM performance has always been that focal tasks have reduced attentional demands, obviating the need for monitoring (though some people may still choose to). Thus, when participants are not monitoring – no task interference is present – there should be no increase in perceived difficulty. Although difficulty ratings for participants in the focal conditions *did* significantly increase, the effect was reduced relative to nonfocal conditions in both Experiment 2 and 3. Still, to verify whether or not focal participants who unnecessarily decided to monitor were responsible for this increase in difficulty, we conducted exploratory analyses conditionalized on whether or not people were considered to be monitoring. If a participant's task interference was less than or equal to zero, they were not considered to be monitoring, but if the value was greater than zero they were considered to be monitoring (irrespective to focality). Two paired-samples *t*-tests found no significant increase in difficulty ratings in either experiment for participants who were not monitoring (Experiment 1: $t(34) = -1.22$, $p = .23$, Cohen's $d = .21$; Experiment 2: $t(31) = -1.64$, $p = .11$, Cohen's $d = .29$). Thus, for participants who showed no task interference, the vast majority of whom had a focal PM task, there was no corresponding increase in difficulty, further supporting a multiprocess perspective.

Incorporating these results with the bulk of the literature, the field now has more direct evidence for what was reasonably assumed in the past: Task interference is associated with perceived increases in difficulty; therefore, when participants report that they use a target-searching strategy (e.g., as they did in Anderson et al., 2018), we can be confident that this behavior falls under the large umbrella of 'cognitively demanding processes'. Now, converging evidence from numerous sources makes it unlikely that self-reported monitoring—for example, participants often report rehearsing the intention or searching for features relevant to the PM target—is epiphenomenal. All indicators assessed so far, self-reported or objectively observed, point toward monitoring (and not delay) as the theoretical construct responsible for task interference.

One concern that future researchers should be aware of, however, is that the difficulty probe procedure was not intended for use beyond addressing a particular theoretical contention; we should acknowledge that the effect sizes observed using this measure were small, and the reliability was likely quite poor, with only four probes per block.[12] Therefore, future work interested in difficulty perceptions should ideally increase the number of probes considerably. However, adding more probes to the experiment would likely be disruptive to the participant, and could compromise the validity of the PM experiment itself. Thus, a truly ideal scenario would be to maintain a similar ratio of probes to trials, and have participants return for multiple sessions (see Strickland et al., 2018 for use of this multi-session methodology). In this way, the integrity of the experiment could be maintained while simultaneously increasing the reliability of the difficulty probes. Further, the validity of the measure could be improved by looking for correlations with more objective physiological measures of difficulty, such as pupil dilation, galvanic skin response, and heart rate. Pupil dilation, in particular, would be interesting, as it is often used as a direct measure of attentional control.

Based on the findings of this study, as well as previous research using a variety of methods to explicitly evaluate the relative strengths of multiprocess and PMDC theories (Anderson & McDaniel, 2019; Anderson et al. 2018, 2019), we encourage PM researchers to favor a multiprocess perspective, pending future work from both theoretical camps. As it stands currently, the bulk of the research supporting PMDC relies on model comparison as a purported litmus test—and they invariably show that the PMDC model fits the data best and fits the data well. Yet, such endeavors require that the models are all appropriately specified, that all relevant models are included, and that the model parameters are psychologically validated as reflecting particular behaviors. Undeniably, the architects of the PMDC model have gone to great lengths satisfying the statistical requirements and creating an impressive model that fits the data well.

[12] To get at the reliability issue, we calculated the average correlation between difficulty probes in the control block (because the control block is likely less contaminated by experimental demands). This correlation was $r = .63$ after transforming from an average Fisher's z score.

However, we also know that the accumulator model parameters have *not* been uniquely linked to PM-related behaviors, instead displaying considerable covariance (Anderson et al., 2018). Specifically, Anderson et al. observed changes in drift-rates, nondecision time, and thresholds when a PM task was added to the OT, and this was true even in conditions designed to isolate delay or monitoring behaviors. Thus, strong inferences about PM based on these model parameters should be avoided, pending more conclusive validation. Yet, even assuming that the other considerations have been reasonably met, the theoretical architecture of the model is contradicted by the bulk of the literature, including this study, and thus cannot be considered valid. The use of accumulator models in PM is an exciting new methodological foray, but if the inferential conclusions such models point toward are simply not sensible, then serious consideration should be given to either abandoning or overhauling that model. At the very least, given the available evidence, PMDC theory should be taken with a grain of salt, and certainly not preferred over a multiprocess view.

References

Abney, D. H., McBride, D. M., & Petrella, S. N. (2013). Interactive effects in transfer-appropriate processing for event-based prospective memory: The roles of effort, OT, and PM cue properties. *Memory & Cognition, 41*, 1032-1045. doi:10.3758/s13421-013-0324-7.

Anderson, F. T., & Einstein, G. O. (2017). The fate of completed intentions. *Memory*, *25*, 467-480.

Anderson, F. T., & McDaniel, M. A. (2019). Retrieval in prospective memory: Multiple processes or just delay?. *Quarterly Journal of Experimental Psychology*, *72*, 2197-2207.

Anderson, F. T., McDaniel, M. A., & Einstein, G. O. (2017). Remembering to remember: An examination of the cognitive processes underlying prospective memory. In J. H. Byrne (Ed.) *Learning and memory: A comprehensive reference 2E* (pp. 451-463). Oxford, UK: Elsevier.

Anderson, F. T., Rummel, J., & McDaniel, M. A. (2018). Proceeding with care for successful prospective memory: Do we delay ongoing responding or actively monitor for cues? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1036-1050.

Anderson, F. T., Strube, M. J., & McDaniel, M. A. (2019). Toward a better understanding of costs in prospective memory: A meta-analytic review. *Psychological Bulletin*, *145*, 1053.

Ball, B. H., & Aschenbrenner, A. J. (2018). The importance of age-related differences in prospective memory: Evidence from diffusion model analyses. *Psychonomic Bulletin & Review*, *25*, 1114-1122. doi:10.3758/s13423-017-1318-4

Ball, B. H., & Brewer, G. A. (2018). Proactive control processes in event-based prospective memory: Evidence from intraindividual variability and ex-Gaussian analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 793.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*, 553-573.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445-459.

Boywitt, C. D., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition*, *40*, 70-82.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. A. Conway, C. Jarrold, M. J. Kane, J. N. Towse, A. A. Conway, C. Jarrold, ... J. N. Towse (Eds.), *Variation in working memory* (pp. 76-106). New York, NY, US: Oxford University Press.

Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition*, *38*, 304-311.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153-178.

Cohen, A. L. (2013). Attentional decoupling while pursuing intentions: A form of mind wandering?. *Frontiers in Psychology*, *4*, 693. doi:10.3389/fpsyg.2013.00693

Cohen, A. L., Jaudas, A., & Gollwitzer, P. M. (2008). Number of cues influences the cost of remembering to remember. Memory & cognition, 36(1), 149-156.

Cona, G., Bisiacchi, P. S., Sartori, G., & Scarpazza, C. (2016). Effects of cue focality on the neural mechanisms of prospective memory: A meta-analysis of neuroimaging studies. *Scientific Reports*, *6*, 1-13.

Conte, A. M., & McBride, D. M. (2018). Comparing time-based and event-based prospective memory over short delays. *Memory*, *26*, 936-945.

Cook, G. I., Rummel, J., & Dummel, S. (2015). Toward an understanding of motivational influences on prospective memory using value-added intentions. *Frontiers in Human Neuroscience*, *9*, 278.

Dismukes, R. K. (2012). Prospective memory in workplace and everyday situations. *Current Directions in Psychological Science*, *21*, 215-220.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 717.

Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: Multiple retrieval processes. *Current Directions in Psychological Science*, *14*, 286-290.

Einstein, G. O., McDaniel, M. A., Thomas, R., Mayfield, S., Shank, H., Morrisette, N., & Breneiser, J. (2005). Multiple processes in prospective memory retrieval: Factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General, 134*, 327.

Guynn, M. J. (2003). A two-process model of strategic monitoring in event-based prospective memory: Activation/retrieval mode and checking. *International Journal of Psychology, 38*, 245-256.

Guynn, M. J., McDaniel, M. A., & Einstein, G. O. (1998). Prospective memory: When reminders fail. *Memory & Cognition, 26*, 287-298.

Harrison, T. L., & Einstein, G. O. (2010). Prospective memory: Are preparatory attentional processes necessary for a single focal cue?. *Memory & Cognition, 38*, 860-867.

Harrison, T. L., Mullet, H. G., Whiffen, K. N., Ousterhout, H., & Einstein, G. O. (2014). Prospective memory: Effects of divided attention on spontaneous retrieval. *Memory & Cognition, 42*, 212-224.

Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory task interference. *Psychological Review, 122*, 376.

Henry, J. D., Rendell, P. G., Phillips, L. H., Dunlop, L., & Kliegel, M. (2012). Prospective memory reminders: A laboratory investigation of initiation source and age effects. *Quarterly Journal of Experimental Psychology, 65*, 1274-1287.

Horn, S. S., & Bayen, U. J. (2015). Modeling criterion shifts and target checking in prospective memory monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 95.

Horn, S. S., Bayen, U. J., & Smith, R. E. (2011). What can the diffusion model tell us about prospective memory?. *Canadian Journal of Experimental Psychology, 65*, 69.

Horn, S. S., Bayen, U. J., & Smith, R. E. (2013). Adult age differences in interference from a prospective-memory task: A diffusion model analysis. *Psychonomic Bulletin & Review, 20*, 1266-1273.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*, 394-412.

Jeong, J. M., & Cranney, J. (2009). Motivation, depression, and naturalistic time-based prospective remembering. *Memory*, *17*, 732-741.

Kliegel, M., Jäger, T., & Phillips, L. H. (2008). Adult age differences in event-based prospective memory: A meta-analysis on the role of focal versus nonfocal cues. *Psychology and Aging*, *23*, 203-208.

Kliegel, M., Martin, M., McDaniel, M., & Einstein, G. (2004). Importance effects on performance in event-based prospective memory tasks. *Memory*, *12*, 553-561.

Krishnan, H., & Shapiro, S. (1999). Prospective and retrospective memory for intentions: A two-component approach. *Journal of Consumer Psychology*, *8*, 141-166.

Leung, S. O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research*, *37*, 412-421.

Loft, S., & Humphreys, M. S. (2012). Enhanced recognition of words previously presented in a task with nonfocal prospective memory requirements. *Psychonomic Bulletin & Review*, *19*, 1142-1147.

Loft, S., Kearney, R., & Remington, R. W. (2008). Is task interference in event-based prospective memory dependent on cue presentation?. *Memory & Cognition, 36*, 139-148.

Loft, S., & Remington, R. W. (2013). Wait a second: Brief delays in responding reduce focality effects in event-based prospective memory. *The Quarterly Journal of Experimental Psychology, 66*, 1432-1447.

Loft, S., & Yeo, G. (2007). An investigation into the resource requirements of event-based prospective memory. *Memory & Cognition*, *35*, 263-274.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology*, *14*(SpecIssue), S127-S144. doi:10.1002/acp.775

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field*. Thousand Oaks, CA: Sage.

McDaniel, M. A., LaMontagne, P., Beck, S. M., Scullin, M. K., & Braver, T. S. (2013). Dissociable neural routes to successful prospective memory. *Psychological Science, 24*, 1791-1800. doi:10.1177/0956797613481233

McDaniel, M. A., Umanath, S., Einstein, G. O., & Waldum, E. R. (2015). Dual pathways to prospective remembering. *Frontiers in Human Neuroscience*, *9*, 392.

McBride, D. M., Beckner, J. K., & Abney, D. H. (2011). Effects of delay of prospective memory cues in an ongoing task on prospective memory task performance. *Memory & Cognition*, *39*, 1222-1231.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299-320.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347-356.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*, 260-281.

Reese, C. M., & Cherry, K. E. (2002). The effects of age, ability, and memory monitoring on prospective memory task performance. *Aging, Neuropsychology, and Cognition*, *9*, 98-113.

Rummel, J., Kuhlmann, B. G., & Touron, D. R. (2013). Performance predictions affect attentional processes of event-based prospective memory. *Consciousness & Cognition, 22*, 729-741. doi: 10.1016/j.concog.2013.04.012

Rummel, J., Smeekens, B. A., & Kane, M. J. (2017). Dealing with prospective memory demands while performing an OT: Shared processing, increased on-task focus, or both?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1047.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*, 414-429.

Scullin, M. K., McDaniel, M. A., & Einstein, G. O. (2010a). Control of cost in prospective memory: Evidence for spontaneous retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 190.

Scullin, M. K., McDaniel, M. A., & Shelton, J. T. (2013). The dynamic multiprocess framework: Evidence from prospective memory with contextual variability. *Cognitive Psychology*, *67*, 55-71.

Scullin, M. K., McDaniel, M. A., Shelton, J. T., & Lee, J. H. (2010b). Focal/nonfocal cue effects in prospective memory: Monitoring difficulty or different retrieval processes?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 736.

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 347.

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 756.

Strickland, L., Heathcote, A., Remington, R. W., & Loft, S. (2017). Accumulating evidence about what prospective memory task interference actually reveal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1616.

Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: A theory of decision control in event-based prospective memory. *Psychological Review, 21*, 03.

Taylor, R. S., Marsh, R., Hicks, J., & Hancock, T. (2004). The influence of partial-match cues on event-based prospective memory. *Memory, 12*, 203-213.

Uttl, B. (2008). Transparent meta-analysis of prospective memory and aging, *PLoS ONE, 3*: e1568.

Uttl, B. (2011). Transparent meta-analysis: Does aging spare prospective memory with focal vs. nonfocal cues? *PLoS ONE, 6*: e16618.

Walter, S., & Meier, B. (2014). How important is importance for prospective memory? A review. *Frontiers in Psychology*, *5*, 657.

West, R., & Bowry, R. (2005). Effects of aging and working memory demands on prospective memory. *Psychophysiology, 42*, 698-712.

West, R., Krompinger, J., & Bowry, R. (2005). Disruptions of preparatory attention contribute to failures of prospective memory. *Psychonomic Bulletin & Review*, *12*, 502-507.

Appendix

Counterbalancing Effects on Response Times in Experiment 1

Motivation counterbalancing order significantly interacted with motivation, $F(1,114) = 11.41$, $p = .001$, $MSE = 1.43$, $\eta_p^2 = .09$, such that the motivation effect (longer response times when motivated) was reduced when the motivated block came second. Additionally, *N*-back counterbalancing order significantly interacted with OT load, $F(1,114) = 100.70$, $p < .001$, $MSE = 2.16$, $\eta_p^2 = .47$. Presumably due to practice effects, response times when the 1-back task came first were statistically equivalent to those obtained for the 2-back task. However, when the 2-back task came first, response times were slower for the 2-back relative to the 1-back task. These two factors further interacted with cue number, $F(1,114) = 8.33$, $p = .005$, $MSE = 2.16$, $\eta_p^2 = .07$, revealing that the difference in response times between the 1-back and 2-back task, only when the 2-back task came first, was greater when participants were given three cues than when they were given one. Motivation counterbalancing, OT load, and motivation also significantly interacted, $F(1,114) = 7.12$, $p = .009$, $MSE = .59$, $\eta_p^2 = .06$. The interaction was reflected in a larger response time difference between the 1-back and 2-back task (2-back being slower) in motivated blocks when motivation came first, and a larger difference between the 1-back and 2-back task in non-motivated blocks when motivation came second. Finally, there was a four-way interaction between motivation counterbalancing, *N*-back counterbalancing, OT load, and motivation, $F(1,114) = 17.48$, $p < .001$, $MSE = .59$, $\eta_p^2 = .13$, which is difficult to interpret, but seems to be some combination of the aforementioned effects.